The Psychology of Common Knowledge: Coordination, Indirect Speech, and Self-Conscious Emotions

The Psychology of Common Knowledge: Coordination, Indirect Speech, and Self-conscious Emotions

A dissertation presented

by

Kyle Andrew Thomas

to

The *Department of Psychology*

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

May 2015

Dissertation Advisor: Professor Steven Pinker Kyle Andrew Thomas

The Psychology of Common Knowledge: Coordination, Indirect Speech, and Self-conscious Emotions

ABSTRACT

The way humans cooperate is unparalleled in the animal kingdom, and coordination plays an important role in human cooperation. *Common knowledge*—an infinite recursion of shared mental states, such that A knows X, A knows that B knows X, A knows that B knows that A knows X, ad infinitum—is strategically important in facilitating coordination. Common knowledge has also played an important theoretical role in many fields, and has been invoked to explain a staggering diversity of social phenomena. However, no previous empirical work has directly explored the psychology of common knowledge.

Paper 1 demonstrates that people represent common knowledge, distinguish it from lower levels of shared knowledge (e.g., A knows that B knows X, but nothing more), and that common knowledge facilitates coordination for mutual benefits. The paper reports results from four experiments in which groups of participants interacted in coordination games, with varying levels of knowledge and payoffs. Results showed that common knowledge facilitates coordination, and thus provides an important proof of concept.

Paper 2 provides support from a large dyadic psychophysiology study for a recently proposed theory of strategic indirect speech, in which common knowledge plays a central role. Participants' affective reactions to different types of illicit propositions were consistent with

predictions from the theory, as were their responses to survey questions that asked what they would tell their friends about the propositions. By supporting the strategic theory of indirect speech, these results provide indirect evidence that common knowledge plays an important role in explaining certain kinds of indirect speech.

Paper 3 provides evidence from two experiments that the self-conscious emotions of embarrassment, shame, and guilt are sensitive to the distinction between common knowledge and lower levels of shared knowledge. In the first experiment, participants read fictional scenarios that might induce these emotions, and reported that they would feel them more strongly if a transgression was common knowledge than if it was merely shared knowledge. In the second experiment, participants performed a karaoke song for a panel of judges, and reported higher levels of embarrassment when their performance was common knowledge than shared knowledge.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Working on my PhD at Harvard has been the greatest opportunity and experience of my life, and I will be forever grateful that I was given such an opportunity. First and foremost, I would like to thank my advisor Steven Pinker for offering me this amazing opportunity, and for all of the incredible guidance and support he has provided along the way. I could not imagine a better mentor. I am also greatly indebted to other faculty members that encouraged, inspired, and supported me in various ways: Wendy Mendes, Jim Sidanius, Max Krasnow, Marc Hauser, Stephen Kosslyn, George Alvarez, Martin Nowak, and Joshua Greene.

I would also like to thank some of the many mentors who have provided invaluable guidance and support throughout my journey: Greg Bryant, Peter DeScioli, Omar Haque, Christopher Oveis, Elizabeth Page-Gould, Jeremy Jamieson, Adam Waytz, Rob Kurzban, Carlos Navarette, Samuel Moulton, Karim Kassam, Clark Barrett, Dave Rand, Rebecca Sutherland, Geoffrey Miller, Bruce Bridgeman, and James Lee. Of course, one of the best aspects of being in the department has been the outstanding caliber of the peers I have met, who have greatly influenced me, and that I am lucky enough to call my friends: Jennifer Sheehy-Skeffington, Nour Kteily, Adrian Ward, Moshe Hoffman, Brendan Gaesser, Jason Nemirow, Arin Tuerk, Samuel Mehr, Steve Lehr, Juan Manuel Contreras, Robin Bergh, Kathy Gerlach, Christopher Nocera, Daniel Rosenbloom, Joshua Hartshorne, Stefano Anzelotti, Katie McAuliffe, and Alek Chakroff.

Most importantly, I would like to thank my wife, mom, dad, and brothers, without whom none of this would have been possible.

INTRODUCTION

The way humans cooperate is unparalleled in the animal kingdom. Coordination plays a central role in human cooperation, such as in organizing social events, agreeing on common currencies, norms, and communication systems, or bringing down a dictatorial leader. Coordination can offer many benefits, but it is also potentially costly, from the awkwardness of a dead party, to lost investments in currencies, norms, or signals that others don't recognize, to brutal repression at a poorly attended protest. Psychologists have largely overlooked coordination, despite its ubiquity throughout human social life. Yet, it reveals important unanswered questions about how we represent other people's mental states, which are critical to understanding how cooperation evolved and works.

This dissertation provides the first direct evidence that humans have psychological mechanisms designed for one such mental state—*common knowledge*—which is defined as an infinite recursion of shared knowledge levels, such that A knows X, A knows that B knows X, A knows that B knows that A knows X, ad infinitum. For decades, the concept of common knowledge has played an important theoretical role in disciplines that span practically the entire range of the social sciences—including economics, sociology, linguistics, legal studies, political science, history, psychology, and anthropology—as well as in game theory, philosophy, and computer science. It is thus unsurprising that common knowledge has been invoked to explain a staggering diversity of seemingly unrelated phenomena: from currency valuation and market bubbles, to pluralistic ignorance and scandals, direct reference and indirect speech, laws and norms, international diplomacy and bargaining, coordination and cooperation, rituals and social constructions, and even how words acquire meaning and databases are networked (for a review,

see Thomas, DeScioli, Haque, & Pinker, 2014). What is surprising is that despite this breadth, no previous research has directly investigated the psychology of common knowledge. The studies presented here begin to fill this gap.

All of the phenomena listed above involve situations in which two or more agents (or even computers) can coordinate their actions for mutual benefit, and game-theoretic analyses of *coordination games* demonstrate how common knowledge is inexorably linked to coordination. Coordination games capture situations in which two or more people can benefit by coordinating their behavior, but there is more than one way to do so, and coordination failure is costly. For example, two individuals may both benefit by meeting up for lunch at one restaurant or another, but each would rather eat at home than incur the cost of an expensive meal without the benefit of the other person's company. Because each person's best option depends on what their partner does, one should do whatever they expect their partner to do, which is dependent on what they expect their partner to expect them to do, and thus on what they expect their partner to expect them to expect, and so on. Common knowledge facilitates coordination by focusing all players' attention on a singular outcome, so that all of these interdependent expectations line up. Furthermore, it can easily be demonstrated that coordination games are the *only* context in which common knowledge affects strategic behavior[1].

This link provides some simple guidelines and a roadmap for psychological research:

1. Common knowledge should affect any psychological domain involving coordination;
2. Common knowledge should only affect behavior in contexts involving coordination;

[1] More precisely, recursive shared beliefs only affect play within games that have multiple equilibria (i.e., common knowledge should only cause agents to shift from one strategy to another in coordination games). It should be noted that common knowledge plays an important role in other areas, such as in preventing speculative trading or allowing for backwards induction, but in these areas it's simply a static assumption, as opposed to dynamically affecting whether players choose one strategy over another.

3. And, if common knowledge is shown to affect behavior in some domain, this is diagnostic of some underlying coordination problem, which can reveal and/or clarify features and functions of psychological mechanisms related to the domain.

These three insights are both simple and profound. They generate numerous testable predictions about the psychology of common knowledge, some of which are explored in Papers 1 and 2. And, Papers 2 and 3 show how common knowledge effects can help elucidate the structure and function of other psychological mechanisms, by revealing design features for coordination.

A Brief History of Common Knowledge: The Mathematics of Coordination

The concept of common knowledge has been around at least since David Hume (1739-1740/1969), but its modern treatment begins with Thomas Schelling (1960). Schelling introduced common knowledge to explain how people could coordinate their behavior if one of many possible options is made uniquely salient to everyone. The concept was then independently formalized by a sociologist, a philosopher, and a game theorist, each of whom developed similar frameworks to address very different problems. It's widely believed that David Lewis (1969) provided the first formalization of common knowledge—as an infinite hierarchy of shared knowledge states—which he developed to explain how words acquire meaning (a coordination problem in which everyone must converge on the same definitions for arbitrary symbols, e.g., "cat" means *cat* and "dog" means *dog*, and not vice versa). However, this was actually predated by Morris Friedell (1967), who independently developed the same formalization in an obscure sociology paper that also presaged the role of common knowledge in a number of the phenomena listed above, such as pluralistic ignorance. Although it came a bit later, it seems that Robert Aumann's (1976) set-theoretic formalization was also developed independently[2], in a line of

[2] See http://www.ma.huji.ac.il/raumann/documents/epistemiclogicfinalversion.pdf

work on how subjective beliefs can affect coordination. Aumann's formalization models two agents who use public information to systematically eliminate possible worlds until both converge on the same real world, which is then commonly known.

Two later papers, both published in the same year, examined what happens when different assumptions about common knowledge are relaxed. Rubinstein (1989) loosened the "common" assumption to evaluate what happens when recursive levels of shared knowledge are finite, and found that even arbitrarily high levels are insufficient for two agents to shift from one coordination option to a better one. Monderer and Samet (1989) instead loosened the "knowledge" assumption to explore what happens when beliefs are less than 100% certain. They replaced common knowledge with *common p-belief*, where *p* represents the probability that two individuals believe something to be true, and demonstrated that as long as *p* is sufficiently high the dynamics of coordination and common knowledge still apply (i.e., common p-belief is sufficient to switch from one coordination option to another). They reconciled their results with Rubinstein's by suggesting that "...a complete analysis cannot be carried out on the basis of truncated hierarchies [of shared knowledge levels, because] two situations in which truncation takes place after the same number of [knowledge levels] may differ dramatically because different beliefs are held beyond the truncation point" (p. 189). More recent work confirms that common p-belief provides a general solution, which can account for Rubinstein's results and be used in games with more than two people (Dalkirin, Hoffman, Paturi, Ricketts, Vattani, 2012).

From the Mathematics to the Psychology

What does all this have to do with psychology? Set-theoretic formalizations and common p-beliefs may seem far from the purview of psychology, but are crucial for understanding how people might identify, assess, and condition their behavior on different kinds of events and

knowledge states. First, common p-belief shows how real people living in an uncertain and noisy world can actually attain something akin to common knowledge, and demonstrates how insights from models that assume complete certainty can still apply to real-world strategic behavior. Second, the models suggest that beliefs must somehow be assessed as sufficiently likely in order to affect coordination, which presents important psychological questions about how people assign subjective likelihoods to shared beliefs and what constitutes "sufficiently likely". Third, although set-theoretic formalizations seem to provide implausible descriptions of actual psychological mechanisms, by specifying what kinds of events, observations, or data might produce relatively certain shared beliefs, they can help define the input criteria that more plausible psychological theories must account for. Lastly, the mathematical models demonstrate that the key issue isn't whether something is 100% certain, or whether people can represent an infinite level of recursive beliefs, but whether certainty decays with increasing levels of recursive beliefs. This last point is not directly explored here, but it plays a critical role in the theory of indirect speech tested in Paper 2, and offers an intriguing avenue for future research.

Cognitive scientists have largely converged on the view that the brain is a computational organ that evolved to process information, and that the most efficient approach to understanding the mind is to reverse engineer its computational mechanisms. According to this view, experimental psychologists should begin with a task analysis to precisely define a problem and the engineering specifications required to solve it, come up with plausible information-processing mechanisms that can accomplish the task, and then translate these hypothesized mechanisms into testable empirical predictions (see Tooby & Cosmides, 2005; Marr, 1982). A proper task analysis requires "some independently established finding of science or mathematics to show that some mechanism can attain some goal in some environment" (Pinker, 2000, p. 441).

Game theory, the mathematics of strategic social interaction, provides the best source for task analyses for most social cognition problems, which cannot be defined by more concrete physical properties (as opposed to how the physics of optics and lenses can be used to reverse-engineer visual systems, for example). The mathematical models outlined above thus help define the properties that information-processing mechanisms must have to solve coordination problems, and also help clarify and delineate exactly how and when common knowledge might affect various domains of human social life. Of course, empirical research is required to establish the precise nature of the actual psychological mechanisms for common knowledge.

Establishing an Empirical Foundation for the Psychology of Common Knowledge

Common Knowledge and Coordination. Paper 1 provides the first direct evidence that humans have cognitive adaptations for common knowledge, and that people use different knowledge states strategically when deciding whether they should try and coordinate with others for mutual benefits. In four experiments, participants interacted in stylized coordination games with real monetary payoffs. In the interactions, participants could earn more by working together than by working alone, but this required all participants in a group of either two or four people to make the same choice; otherwise, participants who decided to work together earned nothing. Participants could thus achieve higher earnings by deciding to work together, but choosing this option also carried the risk of relinquishing a lesser, but guaranteed payoff for working alone. We manipulated both what participants knew about what their partners knew about the prevailing payoffs, as well as the payoff amounts, to test how knowledge levels and the ratio of potential costs and benefits would affect participants' decisions to try and work together.

Across all experiments, the percentage of participants that decided to work together was highest when they had common knowledge, consistent with game-theoretic predictions. We also

found multiple converging lines of evidence suggesting that theory of mind representations for common knowledge are categorically distinct from representations for lower levels of shared knowledge (e.g., A knows that B knows that A knows X, but nothing more). These findings provide the first direct evidence that people have psychological mechanisms for recognizing common knowledge, differentiating it from lower levels of shared knowledge, and strategically conditioning their behavior on the prevailing level of knowledge when trying to coordinate. The paper concludes with a discussion of the many potential psychological applications of common knowledge.

Common Knowledge and Indirect Speech. Paper 2 presents results from a large dyadic psychophysiology experiment that supports a strategic theory of indirect speech in which common knowledge plays a central role. This *strategic speaker theory* proposes that indirect speech offers speakers an optimal way to try and transition from one kind of relationship to another with a like-minded partner (e.g., from friends to lovers), while also preventing unnecessary damage to an existing relationship by preventing common knowledge of any conflict of interests (e.g., one person wants to just be friends, while the other wants to be lovers) (Lee & Pinker, 2010; Pinker, Nowak, & Lee, 2008). The theory is predicated upon abstract models of costs and benefits, and hypothesizes that these costs and benefits are realized psychologically as different affective states. The theory also hypothesizes that indirect speech can help speakers prevent rumors from spreading to third parties.

In the experiment, participants' cardiovascular responses were measured while they acted out two scripted role-playing scenarios. One scenario involved a bribe proposition and the other involved a sexual proposition, each of which was delivered either directly or indirectly. Participants' cardiovascular responses were measured throughout the interactions to provide an

indirect, online, and unobtrusive assessment of their affective reactions to different kinds of propositions to test the affective predictions of the theory. The rumor-prevention hypothesis was tested through survey questions that asked participants to report what was asked in the date scenario, what was meant, and what they might tell a friend.

The cardiovascular results from this experiment were a bit complex, and not completely as predicted, yet they nonetheless seem to support the strategic speaker theory's affective hypotheses. The survey results were much clearer, and unequivocally support the rumor-prevention hypothesis. The paper concludes with a discussion that ties indirect speech to other well-established findings, to suggest that the models employed by the strategic speaker theory apply to many aspects of social life beyond just different speech acts, such as the omission bias in moral cognition and the nature of scandals.

Common Knowledge and Self-conscious Emotions. Paper 3 reports the results of two studies showing that the negative self-conscious emotions of shame, embarrassment, and guilt are sensitive to the distinction between common knowledge and lower levels of shared knowledge. In the first study, participants read stories involving different levels of knowledge, which were designed to evoke either the negative self-conscious emotions or the negative basic emotions of anger or sadness. After each story, participants rated how strongly they would expect to feel different emotions, and how likely they thought they would display associated physical reactions (e.g., blush, hang their head down, clench their teeth, etc.). Results showed that participants expected they would feel more embarrassed, ashamed, and guilty when a transgression is made common knowledge, than when the transgression is merely known by others (e.g., a participant knows other people know about their transgression, but it is not common knowledge), and would similarly be more likely to display physical reactions associated

with the self-conscious emotions (e.g., blush, hang their head down, etc.). In contrast, participants' reports to the basic emotion scenarios showed that simply knowing about the event had the largest effect on their expected reactions, and the difference between shared and common knowledge was minimal.

The second study reveals a similar pattern in actual experiences of embarrassment in the lab. In this study participants performed a karaoke song for a panel of judges, and reported greater embarrassment when their performance was commonly known between them and the judges, than when the participant thought the judges were unaware that the participant knew of their presence. This paper thus provides two lines of converging evidence that the negative self-conscious emotions are sensitive to the distinction between common knowledge, and lower levels of shared knowledge, and concludes with a discussion for why these emotions may have evolved this way.

PAPER 1

THE PSYCHOLOGY OF COORDINATION AND COMMON KNOWLEDGE

Kyle A. Thomas[1], Peter DeScioli[1,2], Omar Sultan Haque[1], Steven Pinker[1]

[1]Department of Psychology, Harvard University

[2]Department of Political Science, Stony Brook University

Abstract

Research on human cooperation has concentrated on the puzzle of altruism, in which one actor incurs a cost to benefit another, and the psychology of reciprocity, which evolved to solve this problem. We examine the complementary puzzle of mutualism, in which actors can benefit each other simultaneously, and the psychology of coordination, which ensures such benefits. Coordination is facilitated by *common knowledge*—the recursive belief state in which A knows X, B knows X, A knows that B knows X, B knows that A knows X, ad infinitum. We test whether people are sensitive to common knowledge when deciding whether to engage in risky coordination. Participants decided between working alone for a certain profit and working together for a potentially higher profit that they would receive only if their partner made the same choice. Results showed that more participants attempted risky coordination when they and their prospective partner had common knowledge of the payoffs (broadcasted over a loudspeaker) than when they had only shared knowledge (conveyed to both by a messenger) or private knowledge (revealed to each partner separately). These results support the hypothesis that people represent common knowledge as a distinct cognitive category that licenses them to coordinate with others for mutual gain. We discuss how this hypothesis can provide a unified explanation for diverse phenomena in human social life, including recursive mentalizing, performative speech acts, public protests, hypocrisy, and self-conscious emotional expressions.

Introduction

A strange and ethereal protest took place in Belarus during the summer of 2011, consisting solely of protesters' phones ringing simultaneously. Police swarmed the event, recorded who was there, and made aggressive arrests (Barry, 2011). What were the protesters trying to accomplish? And why were the police concerned with such a seemingly trivial event?

People interact in a variety of situations in which they need to coordinate their actions to achieve common goals, such as opposing unfair governments, capturing gains in trade, agreeing on the use of standard symbols and protocols, and countless everyday activities such as scheduling meetings, contributing complementary courses to potluck dinners, and carrying two ends of a heavy object. Because it may be costly to engage in a coordinated activity when no one else does so, attempts to coordinate can be risky when it is unclear what other people will do. In repressive regimes a single protestor risks prosecution and violence, a risk which can be mitigated only by overwhelming numbers of people successfully coordinating their actions: If one protestor shows up he gets shot, if a million show up they may send the dictator packing. In these situations, even modest displays of synchrony, such as simultaneous phone rings, can set the stage for larger-scale coordination. However, even when it's clear that other people want to work together, coordination can be a challenge. Exactly how, for instance, do thousands of would-be protestors converge on a single time and place to voice their concerns?

Coordination problems are a subtopic in the psychology of cooperation. Though cooperation has become a burgeoning area in psychology, economics, and evolutionary biology, research and theory have concentrated on the subtype of cooperation that is *altruistic* (in the biological sense): A cooperator confers a benefit on a partner at a cost to himself. Altruistic cooperation has received the lion's share of attention because it raises the evolutionary puzzle of

how a behavior that harms the actor could be selected for. The paradox is often captured in the game-theoretic scenario of the Prisoner's Dilemma, and the challenge to the psychologist is in characterizing the cognitive abilities and emotional motives that allow humans to surmount it. These include the ability to recognize individuals and detect cheaters, and a suite of emotions that police reciprocation, such as sympathy, anger, gratitude, forgiveness, guilt, and trust (Trivers, 1971; Cosmides & Tooby, 1992, 2005b).

Coordination, in contrast, is *mutualistic*: Each cooperator confers a benefit on the other while simultaneously conferring a benefit on himself or herself. Despite this convergence of interests, coordination, too, poses an evolutionary challenge. The challenge is not motivational but epistemological: accurately representing the other actor's state of knowledge. The epistemological problem results from the difficulty of converging on a single solution when more than one is available. For instance, two friends both benefit if they meet at Starbucks, or at Peet's, but for this to happen each friend has to know that the other knows which location they have agreed upon.

If this problem can be resolved, the incentives of the game pose no further obstacle, and can even help guide optimal behavior rather than hinder it (Lewis, 1969; Schelling, 1960; Skyrms, 2004). The paradigm game-theoretic model of a coordination problem is the Stag Hunt, first introduced by Jean-Jacques Rousseau (Rousseau, 1754/1984; Skyrms, 2004). Two hunters set out in the morning either to hunt stag together (a large payoff) or to hunt rabbits separately (a small payoff); a single hunter cannot fell a stag and will return empty-handed (a high opportunity cost). To attain the highest payoff, each hunter must not only know that stag offers higher payoffs, but they must also know that the other hunter knows the payoffs, know that the other hunter knows that they know the payoffs, and so on.

Yet despite this epistemological problem, humans are adept at achieving coordination. Protestors meet up in Tahrir Square at 5pm on Friday, different suppliers produce the parts for a complex product, allied battalions converge on an enemy, diners use the bread plate to the left, coworkers in a building settle on an informal name for a meeting space. Given a long evolutionary history of group living, human cognition may have been shaped by natural selection to solve coordination problems (Tooby & Cosmides, 2010; Tooby, Cosmides, & Price, 2006). If game theorists are correct that common knowledge is needed for coordination, then humans might have cognitive mechanisms for recognizing it.

This paper attempts to begin to redress the imbalance in the literature on the psychology of cooperation by exploring the epistemological challenges and the possible cognitive and motivational adaptations surrounding the problem of mutualistic coordination.[3] We focus on a special kind of representation called *common knowledge* (sometimes called *mutual knowledge* or *common ground*; Clark, 1996; Clark & Marshall, 1981; Lewis, 1969; Pinker, 2007; Rubinstein, 1989; Schelling, 1960; Smith, 1982). Common knowledge is defined as an infinite string of embedded levels of mutual knowledge, i.e., Michael knows *X*; Lisa knows *X*; Michael knows that Lisa knows *X*; Lisa knows that Michael knows *X*; Michael knows that Lisa knows that Michael knows *X*; ad infinitum.

The infinite levels of knowledge required for common knowledge may seem to present a different kind of epistemological problem, namely that a finite mind cannot represent an infinite set of nested propositions. However, people need not represent each level of knowledge explicitly, but could simply represent a recursive formula that entails all levels of knowledge,

[3] A PsychInfo search reveals that in the years 1992-2013, 1,936 papers listed "altruism" as a major subject heading or keyword, whereas only 71 listed "mutualism" (and most of these were for studies of nonhuman animals). There were 400 references to the Prisoners' Dilemma, but only 4 to the Stag Hunt.

such as *Y* = "Everyone knows *X,* and everyone knows *Y*", or even just a single symbol that indicates the state of common knowledge itself (Clark, 1996; Pinker, 2007). This formula or symbol, moreover, can be activated in people's minds by any salient public signal which reliably causes the knowledge, such as a message broadcasted on a loudspeaker: Everyone who receives the signal knows that everyone else has received it, and can deduce that everyone else can deduce that, ad infinitum (Aumann, 1976).

Nor is it necessary that that the commonly entertained propositions be known with absolute certainty, which is often impossible in real-world environments. Coordination may be achieved with the weaker notion of *common belief*, in which two agents each believe that a proposition is likely to be true with probability at least *p*, each believes that the other believes it with probability at least *p*, and so on (Monderer & Samet, 1989). For any situation with a stag-hunt payoff structure, there is a minimum level of *p* whose value depends on the relative advantage of coordination over acting alone, for which it is rational for agents with *common p-belief* to choose to coordinate (Dalkiran et al., 2012). In the rest of this paper, we will use the term *common knowledge* broadly, to include "sufficiently high common *p*-belief".

Common knowledge can be contrasted with what we will refer to as *shared knowledge,* any string of embedded levels of knowledge that falls short of infinity, and with *private knowledge,* knowledge that individuals possess without knowing whether anyone else possesses it. Shared knowledge can be further broken down into distinct levels, such as *second-order* or *secondary* knowledge, in which A knows that B knows X, but nothing else, and *third-order* or *tertiary* knowledge, in which A knows that B knows that A knows X, but nothing else. Common knowledge is intimately connected with the logical problem of coordination; in theory, coordination can be irrational without it. With the help of four experiments in which participants

are given the opportunity to engage in a simple form of economic cooperation, we examine the extent to which people really do depend on common knowledge and other levels of knowledge to achieve coordination.

The Game Theory of Coordination and Common Knowledge

Research in game theory on coordination games shows why shared knowledge may be insufficient for coordination. Technically, coordination games are situations of interdependent decision-making that have multiple equilibria. Conceptually, they are situations in which two or more people each make a decision, with the potential to achieve mutual benefits only if their decisions are consistent (Lewis, 1969; Schelling, 1960). The rendezvous example is a coordination game because both friends benefit from choosing the same location, but that location could be either Starbucks or Peet's. To choose among multiple solutions an individual must take into account what she expects the other actor to do. However, what another actor is likely to do is in turn dependent upon his expectations of what she will do, leading to interdependent expectations that generate an infinite recursion of embedded beliefs.

A classic paper demonstrated the importance of common knowledge for maximizing payoffs from a coordination game: Rubinstein (1989) developed a model which showed that under a specific, restrictive set of assumptions any level of knowledge short of common knowledge is no better than no knowledge at all. Subsequent work has suggested that this conclusion was too strong, and that shared knowledge or less-than-certain beliefs can enable coordination better than private knowledge (Binmore & Samuelson, 2001; Dalkiran et al., 2012; Monderer & Samet, 1989). However, even in these models, common knowledge has a privileged role to play in facilitating coordination, in part because it avoids a second-order coordination problem presented by shared knowledge. With shared knowledge people must decide how many

levels of shared knowledge is enough to attempt coordination: How can individuals be certain that everyone requires the same number of levels of shared knowledge to attempt risky coordination? In short, all of these models demonstrate that common knowledge provides the most effective and reliable path to coordination.

The problem of coordination and common knowledge has been examined by many disciplines, including political science (Ostrom, 1990), philosophy (Hume, 1739-1740/1969; Lewis, 1969; Rousseau, 1754/1984; Skyrms, 2004), economics (Chwe, 2001; Geanakoplos, 1992); linguistics (Clark, 1992, 1996; Smith, 1982), sociology (Willer, Kubuwara, & Macy, 2009; Zuckerman, 2010), legal theory (McAdams & Nadler, 2005), and even computer science (Alberucci & Jäger, 2005; Halpern & Moses, 1990). Yet despite the fact that common knowledge is fundamentally a psychological phenomenon, little is known about the psychology of common knowledge (some notable exceptions include Chaudhuri, Schotter, & Sopher, 2009; Clark, 1996; Lee & Pinker, 2010; Nov & Rafaeli, 2009). We briefly review two literatures, experimental economics and theory of mind, that are indirectly relevant to the phenomenon before outlining our own research questions.

Experimental Economics: Coordination Using Salient Focal Points

A few experiments have examined whether people are better at solving coordination problems than classical game theory suggests. They focus on Schelling's (1960) concept of a *focal point,* an option that stands out from a set of possible choices as uniquely salient, encouraging everyone to converge upon it as a single choice. Schelling suggested that in practice people may rely on focal points to solve coordination problems because they generate common knowledge of a single solution (Schelling, 1960; Sugden, 1995). Mehta, Starmer, and Sugden (1994a, 1994b) examined people's play in coordination games and their ability to converge on

focal points (what they called "Schelling salience"). Participants responded to questions with many possible answers (e.g., "Write down any positive number," and "Name any flower"). In one group, participants were paid to answer any way they wanted. In another, they were paid based on how well their answers matched with those of another randomly chosen participant. Participants were far more successful at coordinating answers when they were trying to do so than when they answered as they wished. This suggests that people can meet the challenge of coordination by identifying it as a problem distinct from the primary demands of a task. Though the finding, by itself, cannot distinguish whether people used shared knowledge or common knowledge to improve their coordination, recent unpublished studies suggest that people really do use common knowledge in these tasks (Bardsley, Mehta, Starmer, & Sugden, 2008; Chartier, Abele, Stasser, & Shriver, 2012).

Theory of Mind Research: Representing Shared Knowledge

Most existing research on knowledge about other people's knowledge falls in the area known as theory of mind, intuitive psychology, mind-reading, or mentalizing, all terms for the mental representation of other people's mental states (Baron-Cohen, 1995; Frith & Frith, 2003; Wimmer & Perner, 1983; for recent reviews, see Apperly & Butterfill, 2009; Saxe & Young, 2013). Developmental psychologists have found that by 6-7 months children are able to use implicit representations of attention, desires, goals, and intentions to guide their behavior (Hamlin, Hallinan, & Woodward, 2008). By fifteen months, children implicitly differentiate their own knowledge from another person's knowledge; for example, infants are surprised when someone seeks out an object in a spot where it was moved when the person was absent (Onishi & Baillargeon, 2005). By 3-5 years, children show an ability to explicitly represent others' mental states in the false-belief task (Callaghan et al., 2005; Wellman, Cross, & Watson, 2001). By 6-7

years, children are able to represent two levels of shared knowledge, as evidenced by their ability to understand that someone else can have false beliefs (Perner & Wimmer, 1985). By adulthood, people can correctly answer questions about fourth-order levels of shared knowledge (e.g., Bob knows that Carol knows that Ted knows that Alice knows X), but they tend to fail questions about fifth-order knowledge (Kinderman, Dunbar, & Bentall, 1998), possibly because this exceeds the capacity of short-term verbal memory (Cowan, 2000).

Although people are capable of representing other people's mental states, they do not always do so effectively. Both adults and children tend to assume that their knowledge is shared by other people. This shortcoming is evident in the well-documented failure of three-year-olds to pass a false-belief task, and is also seen in adults in work on the *curse of knowledge* (Birch & Bloom, 2003, 2007; Camerer, Lowenstein, & Weber, 1989; Keysar, Lin, & Barr, 2003).

Since coordination depends on the ability to anticipate other people's actions, and since people's actions depend on their mental states, one would expect mentalizing ability to facilitate coordination. Indeed, Curry & Jones Chesters (2012) found that people who are better at employing theory of mind are also better at coordinating their answers with other people on questions with many possible responses. Yet, characterizations of theory of mind focus on shared knowledge as the paradigm case, and shared knowledge is in general insufficient to solve coordination problems. Researchers have shown that increasing the salience of shared knowledge in cooperative and competitive environments can lead to more competitive behavior (Epley, Caruso, & Bazerman, 2006; Pierce, Kilduff, Galinsky, & Sivanathan, 2013), and have begun to map the neural correlates of representing shared knowledge (Coricelli & Nagel, 2009). Despite such progress in understanding how people represent shared knowledge, far less is known about how people represent common knowledge and use such representations in coordination.

To appreciate the distinctive role of common knowledge in coordination, recall the example of two friends, Sally and Ann, who are trying to find each other downtown. They previously discussed meeting at Starbuck's or Peet's, but never came to an agreement. Where should Sally go to meet Ann? Sally can represent Ann's knowledge of the two locations, and her desire to meet at the same location, and vice-versa. Yet even if Ann thought it would be best to meet at Starbuck's, and Sally knew that Ann thought so, but Ann worried that Sally thought it would be best to meet at Peet's, Ann might go to Peet's while Sally went to Starbuck's. No matter how many nested levels of shared knowledge Sally represents, she will not know where to look for Ann. Coordination games thus pose a key problem for theory of mind research: How does one read the mind of a mind reader?

The Present Research

In these experiments we examine the cognitive processes underlying coordination. Participants interact with partners in a role-playing scenario that involves a symmetric one-shot coordination game, with payoffs that instantiate a Stag Hunt. In the game, participants must decide either to work alone, which offers a small but certain profit, or to try to work with a partner, which offers the potential to make more money but only if their partner makes the same choice: If a participant chooses to work with a partner but the partner does not, they receive nothing. We test whether people differentiate between shared and common knowledge in making this decision, whether shared knowledge and common knowledge have distinct cognitive representations, and whether people use workarounds to a lack of common knowledge.

The game involves two merchants, a butcher and a baker, who decide each day whether to work independently to sell chicken wings and dinner rolls, respectively, or to work together to sell complete hot dogs, for which they earn more (Figure 1.1). No one will buy just the buns or

just the hot dog meat, so they risk earning nothing if they fail to coordinate their actions. Moreover, participants are told that sometimes the hot dogs can earn them both more money than working independently, but sometimes hot dogs earn less money, so the merchants need common knowledge of higher profits to coordinate. But, their only means of communication with each other is an unreliable messenger boy. (This is a stylized instantiation of the *coordinated attack problem*, see Halpern & Moses, 1990; Rubinstein, 1989.)

		Butcher's options	
		Work Together (Hot Dogs)	Work Alone (Chicken Wings)
Baker's options	Work Together (Hot Dogs)	\$1.10, \$1.10	\$0, \$1.00
	Work Alone (Dinner Rolls)	\$1.00, \$0	\$1.00, \$1.00

Figure 1.1 An interaction between a butcher and a baker in Experiment 1. The baker chooses a row and the butcher chooses a column. The four cells show the payoffs (baker's payoff, butcher's payoff) for each combination of choices. These payoffs generate a coordination game, specifically a *stag hunt game* (Skyrms, 2004), in which one equilibrium is better for both players than another equilibrium. The other three payoff conditions substitute \$2.00, \$5.00, or \$10.00 for the \$1.00 payoff, and \$2.10, \$5.10, or \$10.10 for the \$1.10 payoff.

To appreciate the need for common knowledge in this scenario, consider what happens on a given day that the baker sends a message to the butcher telling him to bring hot dogs. The butcher sends a confirmation to let the baker know he received the message. The baker receives the confirmation, but realizes that the butcher cannot be sure whether the messenger delivered the confirmation. So, the baker sends a confirmation of the confirmation. Upon receipt of this message, the butcher realizes that yet another confirmation is required. In fact, no finite number of successful confirmations can help the hapless merchants because they can never be sure that the most recent confirmation message was delivered by the unreliable messenger boy, and neither knows how many messages might be sufficient for the other merchant to bring his

ingredient for the hot dogs (embodying the second-order coordination problem presented by multiple shared knowledge solutions). Common knowledge is therefore needed to reliably solve the merchants' problem.

To test whether people tacitly appreciate this requirement, we manipulated what they knew about their partner's knowledge about the payoffs—whether knowledge of the payoffs was private, shared, or common. The game-theoretic analysis of coordination suggests the *common knowledge recognition hypothesis*: In coordination environments, people strategically differentiate between shared knowledge and common knowledge, working together more frequently when they have common knowledge of the payoffs than when they have shared or private knowledge.

Alternatively, people may not represent common knowledge as a distinct state. The only major distinction affecting their coordination decisions would then be the difference between private and shared knowledge (as suggested by the theory of mind literature). We call this the *shared knowledge hypothesis*.

Finally, the literature on the curse of knowledge raises the possibility that people do not reliably use either shared or common knowledge to solve coordination problems. If people attribute their own knowledge to other people, then distinctions among levels of knowledge would be irrelevant. This *curse of knowledge hypothesis* predicts that participants will try to work together with the same frequency across all knowledge levels.

Knowledge-level Representation

If people do distinguish between common and shared knowledge, this raises the further question of how they represent the distinction. One possibility is that these knowledge states have a single cognitive format and the distinction between them is simply quantitative, with

common knowledge represented as an upper limit of shared knowledge. Alternatively, shared and common knowledge may have distinct representations, which would make the distinction qualitative and categorical.

These possibilities can potentially be distinguished by the pattern of classification errors people make when reporting their level of knowledge. Research on theory of mind capabilities suggests that shared knowledge becomes more difficult to represent as the levels of knowledge increase. If people entertain a single kind of representation (the *single-representation hypothesis*), then the most errors will be observed in the common knowledge condition (the maximum number of shared knowledge levels), with fewer errors made as the levels of shared knowledge decrease. If, in contrast, common knowledge has its own representation, it need not contain multiple levels of embedded knowledge; it could consist of a single mental symbol which means, "We have common knowledge." Thus, errors will increase only with the number of levels of shared knowledge, whereas errors on common knowledge will be few (similar to the error rate for secondary knowledge, which also requires only a single level of representation). The *distinct-representation hypothesis* makes the further prediction that errors will be systematic, respecting the boundary between the two kinds of knowledge: Different levels of shared knowledge will be mistaken for each other, but not for common knowledge.

Sensitivity to Costs and Benefits

To further characterize the decision processes behind coordination, we vary the game payoffs to test people's sensitivity to costs and benefits. Rational choice theory (e.g., Becker, 1976) models human decisions as expected utility calculations, which in coordination situations involves comparing the payoffs of different decisions scaled by the probabilities of achieving each outcome. According to this approach, people should decide to work together if they

determine that the expected value—the amount earned for successful coordination multiplied by the probability that they think their partner will do the same—is greater than the amount earned for working alone. This *rational actor hypothesis*, in which the assumptions of economic modeling are directly interpreted as a psychological theory, predicts that as the benefits of working together decrease relative to the payoff for working alone, people will be less likely to try to work together, and their decisions will track this ratio in the interaction.

In contrast, a more explicitly psychological approach treats game-theoretic models not as literal theories of decision-making, but as task analyses to help identify possible evolved information-processing mechanisms (Gigerenzer, Todd, & ABC Research Group, 1999). The probabilities associated with another individual's behavioral choices are not directly accessible to a human perceiver, but must be heuristically inferred in real time from available cues. Rather than basing their decisions on potentially indeterminate probabilistic calculations of expected utility, humans may instead use simpler heuristics, which categorically distinguish common, shared, and private knowledge using ecologically typical cues to infer whether another person will attempt to coordinate. This *knowledge-level heuristic hypothesis* predicts that people's decisions will be driven primarily by categorical perceptions of knowledge states, and thus may be relatively insensitive to the costs and benefits of coordination.

Other Motivations for Coordination

Coordination decisions may be influenced by factors other than public knowledge. Since coordination requires elusive knowledge that another person has made the same choice as oneself, people may use their own decision-making processes to simulate how their partners will think and behave (Gallese & Goldman, 1998), particularly when they view themselves as similar to their partner (Mitchell, Macrae, & Banaji, 2006; Tamir & Mitchell, 2013). Indeed, perceived

similarity has been shown to help people solve coordination problems in which coordination requires that both make the same decision, but not when it requires that they make different decisions (Abele, Stasser, & Chartier, 2013). According to this *perceived similarity hypothesis*, the more similar an actor perceives himself or herself to be to a potential coordination partner, the more likely the actor will be to predict that their partner will choose as they do, and thus attempt risky coordination.[4]

People may also be motivated to coordinate by altruistic or other-regarding preferences. Altruistic and reputational motives have been well documented in social psychology and experimental economics (e.g., Haley & Fessler, 2005; Messick & McClintock, 1968; Milinski, Semmann, & Krambeck, 2002; Van Lange, 1999). The Big-Five personality trait of Agreeableness is associated with altruism, pro-sociality, friendliness, heightened self-presentation concerns, and generosity (Goldberg, 1992; Graziano & Tobin, 2002; Roccas, Sagiv, Schwartz, & Knafo, 2002; Sun & Wu, 2012), and has specifically been associated with altruistic motivations towards non-relatives and strangers (Graziano, Habashi, Sheese, & Tobin, 2007). The *altruistic motives hypothesis* predicts that people who are higher on Agreeableness will be more likely to try to work together.

Finally, some people may simply be willing to accept the potential cost of discoordination in the hope that they can earn more money through high-payoff coordination. The Big-Five personality trait of Openness is associated with risk-seeking (Nicholson, Soane, Fenton-O'Creevy, & Willman, 2005), and in particular with the seeking of chances for gains

[4]This is related to the concept of superrationality, in which rational actors decide to cooperate in a Prisoner's Dilemma because they each assume that both they and their partner rationally see the wisdom of mutual cooperation, and know that the other sees it, knows that the other knows that they know that the other sees it, and so on (see Colman, 2003; Fischer, 2009; Hofstadter, 1985). The same logic can be applied to coordination games with symmetrical payoff structures.

(Lauriola & Levin, 2001). The decision to work together is as a social gamble where one can bet a certain payout to win an additional increment of profit. The *risk-seeking hypothesis* predicts that people who are higher on Openness will be more likely to take the bet by trying to work together.

We report four experiments designed to test these hypotheses. In Experiments 1 and 2, we test the effects of knowledge level and payoff structure on coordination decisions involving, respectively, one and three partners. In Experiment 3 we investigate how shared and common knowledge are cognitively represented and test the three social motivation hypotheses. In Experiment 4 we verify that the participants' coordination behavior generalizes to a different fictional context and thus is not an artifact of specific features of the first scenario.

Experiment 1

Experiment 1 implements the butcher-baker coordination game explained above. Each participant interacts with a partner, playing the role of either the butcher or the baker. They read that they could work either alone or with the partner; the amount they could earn for working alone was constant, but the amount earned for working together would vary from day to day and might be less than or greater than the amount they could make by working alone. They were then told that on the day of the actual decision facing them the payoff for working together was greater than the payoff for working alone, and were then given one of four kinds of information about what their partner knew about the payoff, which we varied in a between-subject design.

In the *private knowledge* condition, a participant was told he or she could earn 10 cents more for working with the partner, but were not given information about what the partner knew. In the *secondary knowledge* condition, the participant was told that their partner also knew about this payoff. In the *tertiary knowledge* condition, they were told that their partner knew the payoff

and knew that the participant himself or herself knew the payoff. In the *common knowledge* condition, the payoff was presented as public information, commonly known between the two participants. This information was given to participants in one of two boxes presented on their screen. Private, secondary, and tertiary knowledge information was presented in a private knowledge box that they were told only they could see, while common knowledge information was presented in a public knowledge box that both they and their partner could see.

To see whether coordination decisions were sensitive to costs and benefits, we manipulated, between participants, the amount the participants could earn by working alone and together, yielding four payoff structures (see Figure 1.1): $1.00/$1.10, $2.00/$2.10, $5.00/$5.10, and $10.00/$10.10 (hereafter referred to as the $1, $2, $5, and $10 payoff conditions). We note that these stakes are typical for games run on Mechanical Turk (Amir, Rand, & Gal, 2012), and that previous research has shown that while real cash incentives are critical, the absolute size of the stakes tends not to alter the pattern of results (see Camerer, 2003 and Camerer & Hogarth, 1999 for review). We chose small incremental payoffs for coordination (and thus relatively high opportunity costs for discoordination) to counter the typical demand characteristics of experimental games, which tend to encourage cooperative actions (Pedersen, Kurzban, & McCullough, 2013).

Method

Participants. We used Amazon Mechanical Turk to recruit 1600 participants (100 per condition) from the United States to complete a short study for a small payment. Mechanical Turk presents special challenges to ensuring internal and external validity, since online participants may not understand or engage with a task. Data quality and validity can be enhanced with either ex ante prescreening tasks or ex post exclusion based on comprehension questions.

Research shows that both methods can effectively reduce statistical noise without systematically biasing the results (Berinsky, Margolis, Sances, 2013; Goodman, Cryder, & Cheema, 2012; Horton, Rand, & Zeckhauser, 2011; Oppenheimer, Meyvis, & Davidenko, 2009; Rand, Greene, & Nowak, 2012; Suri & Watts, 2011). However, prescreening with comprehension checks can potentially alter participant behavior (Rand et al., 2012), and may increase self-presentation effects (Clifford & Jerit, 2012). Hence, we used ex post exclusion based on comprehension questions about the game's payoff structure (see Procedure).

The final sample consisted of 1033 participants (58% female) with a mean age of 32.8 years (*SD* = 15.0). This 35% ex post exclusion rate is within the range of rates observed in previous research using multiple comprehension checks with Mechanical Turk samples, which can vary between 25% and 50% (Berinsky et al., 2013; Downs, Holbrook, Sheng, & Cranor, 2010; Goodman et al., 2012; Horton et al., 2011; Rand et al., 2012). Furthermore, researchers that have compared similar procedures in the lab and on Mechanical Turk have observed equivalent exclusion rates (Oppenheimer et al., 2009; Paolacci, Chandler, & Ipeirotis, 2010; Rand et al., 2012).

Procedure. Participants read instructions explaining that they would earn a minimum of 50 cents, which they could augment based on their decisions in their interaction with another participant on Mechanical Turk[5]. They were told that one of them would play a butcher and the other a baker. Each could either work alone for a sure profit (the butcher could make chicken wings, the baker dinner rolls) or attempt to work with their partner, the butcher making hot dogs, the baker the buns. By choosing to work together, they were told, the participant could earn a profit, but *only if the partner also chose to work together*; if either decided to collaborate but the

[5] A sample of the experimental stimuli can be viewed at http://www.pdescioli.com/CK_Materials/CK_Materials.html

partner did not, that person would not earn anything, because one cannot sell a bun without a hot dog or vice versa. Participants then read that they would earn a certain amount ($1, $2, $5, or $10) if they decided to work alone, but that the hot dog price varied from day to day, and thus the earnings for working together might be more than or less than this sure profit. Finally, they read that the information about hot dog earnings might be conveyed to them by a messenger boy (displayed on their screen in a private box they were told only they could see), or by a loudspeaker (displayed on their screen in a public box they were told the other participant could see on his or her screen as well).

The participant then clicked a button to reveal the day's information about the price of hot dogs and hence the potential profit for collaborating; in each case it was ten cents more than each would earn by working alone. In the second between-subjects manipulation, participants received one of the following pieces of information (presented here from the perspective of the baker):

1. *Private knowledge*—In the private box the participant read, "The Messenger Boy has not seen the Butcher today, so he cannot tell you anything about what the Butcher knows." The public box stated that the loudspeaker was silent.
2. *Secondary knowledge*—In the private box the participant read, "The Messenger Boy says he stopped by the butcher shop before coming to your bakery. He tells you that the Butcher knows what today's hot dog price is. However, he says that he forgot to mention to the Butcher that he was coming to see you, so the Butcher is not aware that you know today's hot dog price." The public box stated that the loudspeaker was silent.
3. *Tertiary knowledge*—In the private box the participant read, "The Messenger Boy mentions that he is heading over to the butcher shop, and will let the Butcher know

today's price as well. The Messenger Boy will also tell the Butcher that he just came from your bakery and told you the price. However, the Messenger Boy will not inform the Butcher that he told you he would be heading over there. So, while the Butcher is aware that you know today's price, he is not aware that you know that he knows that." The public box stated that the loudspeaker was silent.

4. *Common knowledge*—In the public box the participant read, "The loudspeaker broadcast the market price of [today's price] (of which you could earn [earnings for working together])." In the private box the participant read, "The messenger boy did not come by. Because the market price was broadcast on the loudspeaker, the Butcher knows [today's price], and he knows that you know this information as well."

The participants then made a decision to work alone or with the partner and indicated the choice with the keyboard. They were then asked to explain how they made the decision, and were given two sets of comprehension questions. The first set was used to (ex post) exclude participants who did not understand the game's payoff structure, and contained three questions about the profits under various combinations of decisions:

1. "If you and the Butcher both chose to make hot dogs then how much money would you earn for the interaction?"
2. "If you chose to make hot dogs but the Butcher did not, then how much money would you earn for the interaction?"
3. If you chose not to make hot dogs, regardless of what the Butcher decided, then how much money would you earn for the interaction?"

The second set was used as a manipulation check, and contained four questions about what they and their partner knew:

1. “Do you know the price of hot dogs today?
2. “Does the Butcher know the price of hot dogs today?”
3. “Does the Butcher know that you know the price of hot dogs today?”
4. “Does the Butcher know that you know that he knows the price of hot dogs today?”

Finally, participants filled out a brief demographic questionnaire, submitted the task, and received the base rate payment for completion. Offline, we randomly paired participants to implement the conditions described in the scenario. To calculate each individual’s payoffs, private-level participants were matched with private-level participants and common-level participants were matched with common-level participants, but secondary-level were matched with private-level participants, and tertiary-level were matched with secondary-level participants. (Participants were paid based on their pairing with lower-level knowledge partners, and not on pairings with higher-level knowledge partners.) This matching procedure is necessary to avoid deception because the shared knowledge instructions specify a partner at a lower knowledge-level (which is required to establish secondary and tertiary knowledge). Participants were paid their game earnings using Mechanical Turk’s “Bonus Payment” feature.

Results and Discussion

Figure 1.2 shows that with all four payoffs, the percentage of participants who tried to work together was significantly affected by their state of knowledge (see also the first row of Table 1.1).

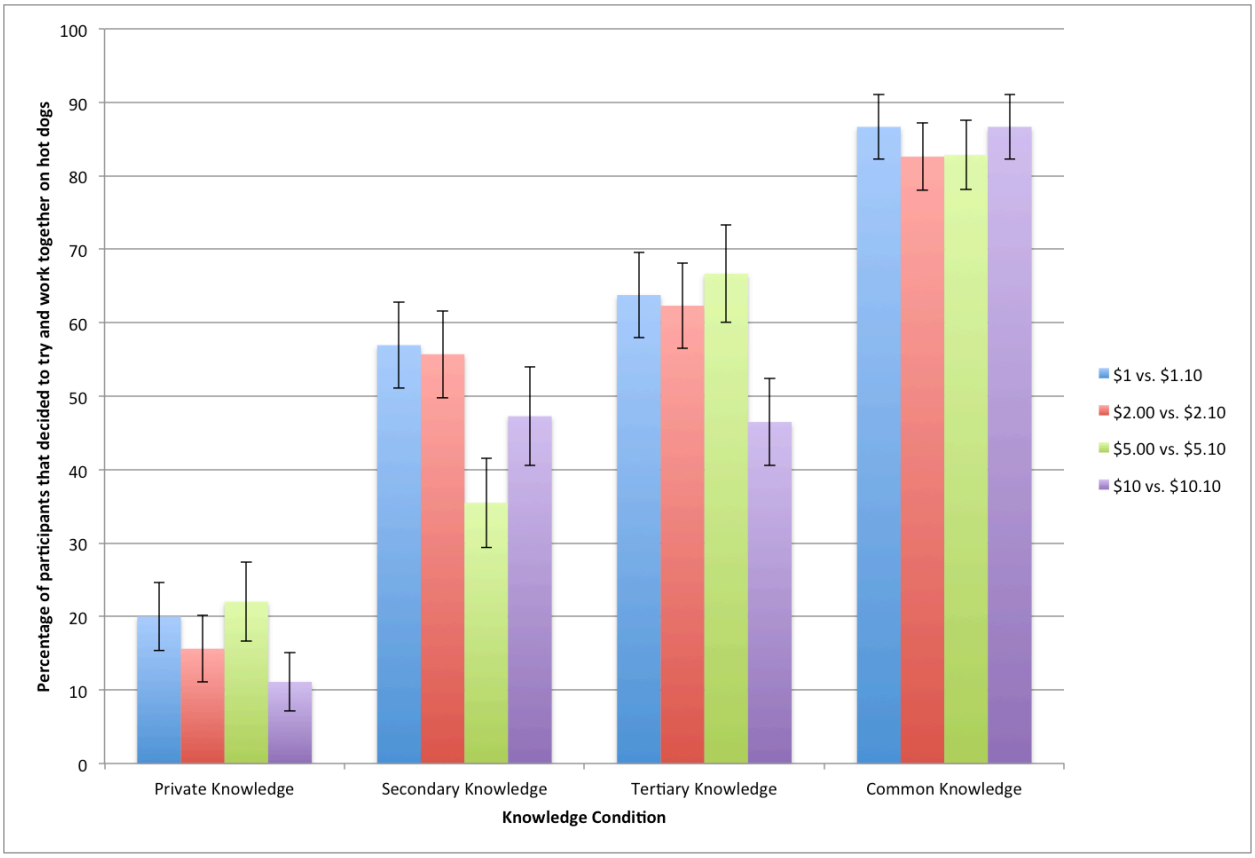

Figure 1.2 Percentage of participants who tried to work together in Experiment 1, organized by knowledge condition and payoff condition. Error bars represent standard error.

Table 1.1 Comparison of knowledge levels in each payoff condition, Experiment 1

Knowledge Levels	$1.00 Payoff			$2.00 Payoff			$5.00 Payoff			$10.00 Payoff		
	χ^2	n	ϕ	χ^2	n	ϕ	χ^2	n	ϕ	χ^2	n	ϕ
All Levels	63.70***	276	.48	62.79***	272	.48	56.71***	236	.49	70.41***	249	.53
Private vs. 2^o	21.26***	147	.38	23.27***	134	.42	2.66	121	.15	19.06***	118	.40
2^o vs. 3^o	0.69	141	.07	0.63	139	.07	10.88***	113	.31	0.01	126	.01
3^o vs. CK	8.84**	129	.26	7.12**	138	.26	4.02*	115	.19	23.05***	131	.42

Note. Chi-square tests for the proportions of subjects who tried to work together across all knowledge levels, and adjacent knowledge levels by payoff condition. We compare private, secondary (2^o), tertiary (3^o), and common knowledge (CK). The comparisons across all knowledge levels have three degrees of freedom, and the comparisons across adjacent knowledge levels have one degree of freedom.
*$p < .05$, **$p < .01$, ***$p < .001$.

Planned comparisons across adjacent knowledge conditions (i.e., private-secondary, secondary-tertiary, and tertiary-common) are shown in Table 1.1 for all payoff conditions. In all four payoff conditions, more participants tried to work together with common knowledge than

with tertiary knowledge. In three out of four payoff conditions, more participants tried to work together with secondary knowledge than with private knowledge (the difference was only marginally significant in the $5 payoff condition). Coordination rates were the same with secondary and tertiary knowledge, except with the $5 payoff, for which the rate with secondary knowledge was anomalously low.

These results are consistent with the Common Knowledge Recognition hypothesis: Participants were more likely to try to work together with common knowledge than with any other state of knowledge. The results were inconsistent with a strong Curse of Knowledge hypothesis, because the likelihood of working together differed across knowledge conditions. In line with the Shared Knowledge hypothesis, few of the participants tried to work together with private knowledge and more tried to work together with secondary knowledge. However, only slightly more participants tried to work together with tertiary knowledge, while far more participants tried to work together with common knowledge (statistics presented in Table 1.1). The pattern is consistent with the hypothesis that people maintain a dual representation in which shared and common knowledge are thought of as qualitatively distinct.

These results were inconsistent with a strict Rational Actor hypothesis because the proportion of participants who decided to try to work together in each knowledge condition varied little across the payoff conditions, even as the ratio between the cost of the forgone profit from working alone to the additional benefit from working together increased tenfold (Private: $\chi^2(3, N = 261) = 3.13$, $p = .373$, $\phi = .11$; Secondary: $\chi^2(3, N = 259) = 7.66$, $p = .054$, $\phi = .17$; Tertiary: $\chi^2(3, N = 260) = 6.81$, $p = .078$, $\phi = .16$; Common: $\chi^2(3, N = 253) = .76$, $p = .859$, $\phi = .05$). A rational actor would expect that as this ratio increases, the other rational actor would take the increased opportunity cost into account, and the probability that they would try to work

together should correspondingly decrease, creating a positive feedback loop that would drive each of them to work alone. The fact that the proportion of people who tried to coordinate with common knowledge was invariant across payoffs contradicts the idea that coordination decisions were based on maximizing the expected payoff, and is instead consistent with the Knowledge-level Heuristic hypothesis.

Another test of the Rational Actor hypothesis may be obtained by examining the actual payouts that the participants would earn given their collective pattern of choices. Inspection of the frequency of coordination attempts in the secondary and tertiary knowledge conditions reveals that this payoff is likely to be low: Participants who decided to try to work together generally failed to coordinate with their partners (and thus relinquished their sure profit from working alone). To assess the overall rationality of these choices we calculated *expected earnings* based on all possible matchups with the other participants (rather than the actual earnings from the matchups we arbitrarily arranged in order to calculate their payments). This consists of the sum of the proportion of participants who chose to work alone, multiplied by the smaller certain payoff, and the proportion of participants that would, on average, successfully work with a partner (when both they and the partner chose to coordinate, which is the product of the proportion of participants that tried to work together and the proportion of potential partners that made the same choice), multiplied by the higher risky payoff. The discoordination payoff, from cases in which they would choose to cooperate but their partner would not, was zero, eliminating this term from the calculation. Figure 1.3a shows that for all payoffs, efficiency was higher with private and common knowledge than with either level of shared knowledge.

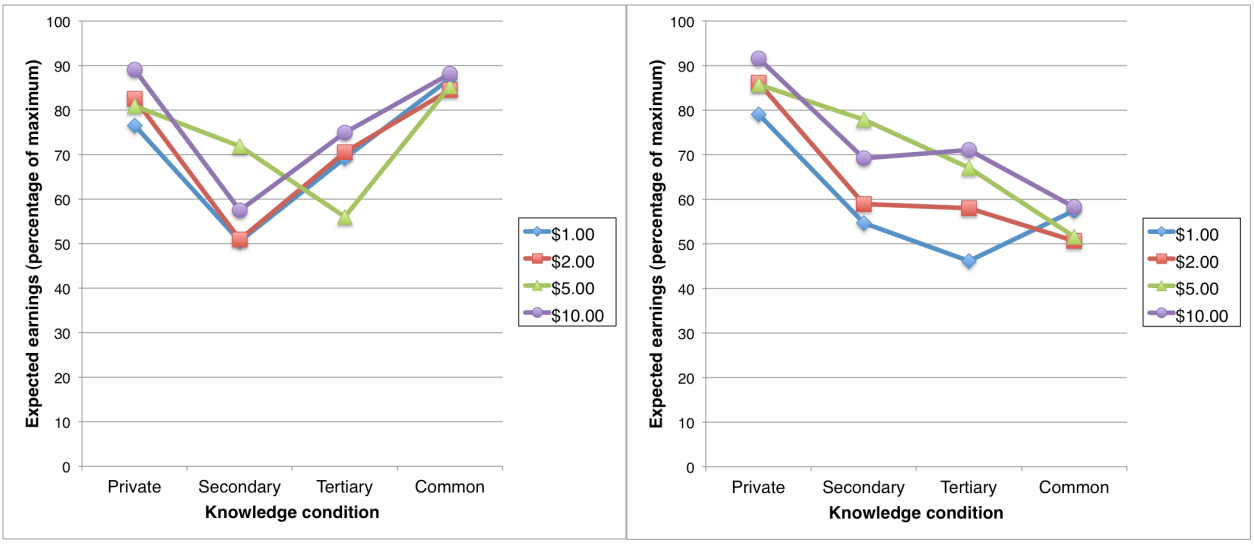

Figure 1.3 Average expected earnings as a percentage of maximum possible earnings for Experiment 1 (*1.3a*, on the left) and Experiment 2 (*1.3b*, on the right) by knowledge condition. We calculated *expected earnings* as the average amount a participant would earn across all possible pairings with the other participants.

In sum, Experiment 1 shows that when people make coordination decisions, they differentiate between private, shared, and common knowledge (though apparently not among different levels of shared knowledge). Moreover, the level of knowledge, and the special appeal of common knowledge, are far more salient to them than the expected value of the options: Increasing the cost-benefit ratio tenfold had no observable impact on their choices.

Experiment 2

How general is the sensitivity to knowledge and insensitivity to payoffs observed in Experiment 1? Presumably, if achieving coordination is difficult enough, and the stakes are high enough, then even with common knowledge people would opt to work alone; as an extreme example, imagine risking a sure payoff of $1000 for working alone for a chance at earning $1001 by coordinating with a million partners. To test the limits of common knowledge as a qualitative coordination heuristic, we designed Experiment 2 as a four-person coordination game in which *all four* partners had to decide to work together to achieve the benefits of coordination.

Coordination on the higher-paying option of working together is far more difficult with four people, because the probability of success is equal to the probability that any one partner decides to work together cubed.

In fact, the perceived probability of successful coordination may fall even faster than that. In addition to common knowledge of the payoffs, coordination also requires individuals to be confident in their partners' rationality. An irrational partner could prefer lower payoffs, choose blindly, or make some other unpredictable choice. With only two players, the chance of an irrational partner might be negligible, but this risk can be greater in larger groups. Since even a single irrational partner can be enough to torpedo coordination in a group, as the number of players goes up the likelihood of discoordination increases rapidly (everyone must be both knowledgeable *and* rational, and believe everyone else is as well). For this reason the cost-benefit structure may become more salient to a participant as the number of other partners increases. Recall that a rational actor may choose to coordinate with less-than-perfect common knowledge (i.e., with common *p*-belief) as long as the probability of the other's belief exceeds a critical value which depends on the relative payoffs: The higher the opportunity cost, the higher that probability must be. Thus we may see a greater sensitivity to payoffs in a coordination game involving more people.

Methods

Participants. As in Experiment 1, 1600 participants were recruited from Amazon Mechanical Turk, evenly distributed across the sixteen combinations of four payoff and four knowledge conditions. After we excluded participants who did not understand the payoffs, the sample consisted of 1150 participants (48% female, $M_{age} = 31.9$, $SD_{age} = 11.1$).

Design and procedure. Participants were told they could work together to make "superburgers," which require a burger, a bun, cheese, and toppings from, respectively, a butcher, a baker, a cheese maker, and a produce vender. One participant was assigned to each of these four roles. As in Experiment 1, each participant also had the option to make a food item on his or her own for a sure profit. Participants were told that they would receive a profit for contributing to superburgers only if *all three* of the other merchants made the same choice, and would receive nothing otherwise. In each of the four knowledge conditions the participant's three partners were said to have the same level of knowledge. This was conveyed with identical instructions to those of Experiment 1, except that "the Butcher" or "the Baker" was replaced with "the other merchants" (so all other merchants were said to have the same level of knowledge). All other aspects of the procedure were the same as in Experiment 1.

Results and Discussion

As in Experiment 1, players' state of knowledge affected their decision to work together (Figure 1.4). Table 1.2 shows that in all payoff conditions, significantly more participants tried to work together with common knowledge than with tertiary knowledge, and in three of the four payoff conditions, significantly more tried to work together with secondary than with private knowledge. In none of the payoff conditions was there a significant difference between secondary and tertiary knowledge. This consistent lack of significant differences between the shared knowledge conditions suggests that people treat secondary and tertiary knowledge similarly.

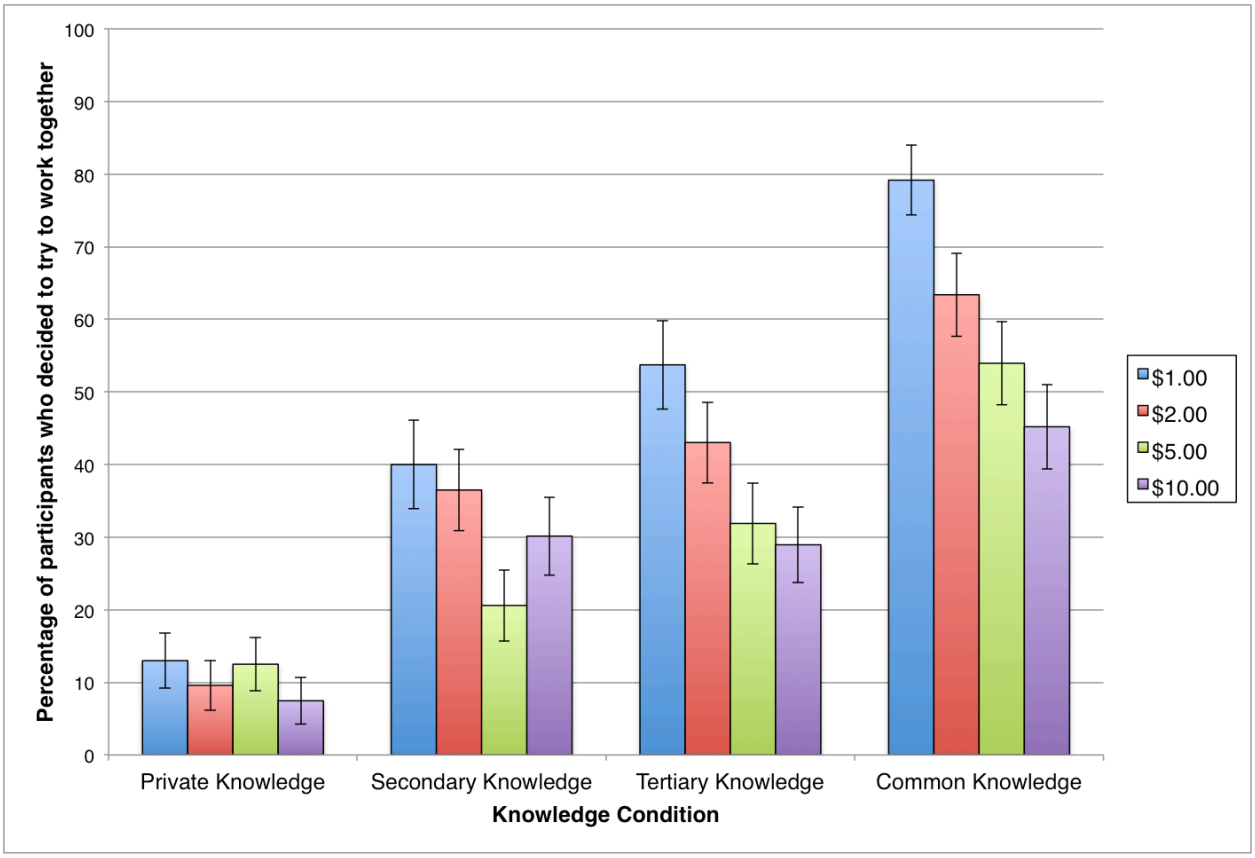

Figure 1.4 Percentage of participants who tried to work together in Experiment 2, organized by knowledge condition and payoff condition. Error bars represent standard error.

Table 1.2 Comparison of knowledge levels in each payoff condition, Experiment 2

Knowledge Levels	$1.00 Payoff			$2.00 Payoff			$5.00 Payoff			$10.00 Payoff		
	χ^2	n	ϕ	χ^2	n	ϕ	χ^2	n	ϕ	χ^2	n	ϕ
All Levels	68.24***	281	.49	45.32***	297	.39	46.36***	283	.41	24.72***	289	.29
Private vs. 2^o	13.59***	142	.31	14.95***	147	.32	1.77	148	.11	11.54***	140	.29
2^o vs. 3^o	2.50	132	.11	0.69	153	.07	2.26	137	.13	0.03	149	.01
3^o vs. CK	10.14**	139	.27	6.21*	150	.20	12.39***	135	.30	4.23*	149	.17

Note. Chi-square tests for the proportions of subjects who tried to work together across all knowledge levels, and adjacent knowledge levels by payoff condition. We compare private, secondary (2^o), tertiary (3^o), and common knowledge (CK). The comparisons across all knowledge levels have three degrees of freedom, and the comparisons across adjacent knowledge levels have one degree of freedom.
*$p < .05$, **$p < .01$, ***$p < .001$.

Unlike Experiment 1, people showed some sensitivity to the payoff structure. In all knowledge conditions, increasing the relative costs of coordination failures brought down coordination rates. This is consistent with the observation that the minimum level of confidence

in common knowledge (i.e., the minimum common *p*-belief) required for rational coordination rises more steeply with opportunity costs when the number of players (and hence the chance that at least one will be ignorant or irrational or both) increases.

Though participants' sensitivity to payoffs was more consistent with the Rational Actor hypothesis than in Experiment 1, another aspect of the results was not. Unlike what we obtained in Experiment 1, the most profitable knowledge condition with all four payoffs was Private Knowledge (Figure 1.3b); Common Knowledge was *less* profitable than Shared Knowledge except with the least costly forgone payoff of $1. This shows an important limit to the advantages that people can obtain from common knowledge. When either the knowledge state or the rationality of all the necessary potential partners is less than perfect, coordination is difficult to achieve and hence poses a high risk of failure. In those cases even high rates of decisions to coordinate may not be enough to consummate successful coordination, and the temptation to coordinate presented by common knowledge can actually *reduce* the coordinators' payoff. Yet more than half of the participants provided with common knowledge still opted for the risky higher payoff.

Experiment 3

Experiment 3 is a replication of one of the payoff conditions from Experiment 1 with additional components that allow us to test how shared and common knowledge are represented, and why people sometimes make what appears to be an irrational decision to cooperate with just shared knowledge.

At least since Miller & Nicely (1955), cognitive psychologists have used confusion matrices to test hypotheses about underlying mental representations, based on the assumption that confusable stimuli are likely to be represented similarly. A similar logic underlies the

memory confusion paradigms commonly used in social psychology to reveal the dimensions of social categorization, such as the "Who said what?" paradigm (e.g., Klauer & Wegener, 1998; Lieberman, Oum, & Kurzban, 2008; Taylor, Fiske, Etcoff, & Ruderman, 1978). In our case, we use errors in responses to our questions about participants' comprehension about the level of knowledge as evidence of whether shared and common knowledge are represented in the same or in qualitatively distinct ways. Unfortunately, in the first two experiments these questions were so easy that almost all participants got them all correct. In this experiment, we made the questions more difficult in three ways: by putting them at the end of the survey, by adding a task before participants answered them, and by concealing the relevant information while they answered the questions (in the first two experiments, this information was visible on the screen).

This experiment tests several additional hypotheses. Recall from the Introduction that there are several reasons that people may choose to coordinate even in the absence of common knowledge. One is that an actor may ascertain that she has similar values and biases to a potential partner, and thus that the partner is likely to assess the situation in the same way that she does, including an assessment of whether she herself is likely to choose to coordinate with the partner. We thus manipulated whether the participants thought they were interacting with a partner who was similar or dissimilar to themselves in age, political orientation, tastes in music, and decision-making style. The other social motivations for coordination consist of personality traits that make the choice inherently appealing, including Agreeableness, which impels people to act in a pro-social manner, and Openness, whose risk-seeking component may impel people to gamble for a big payoff rather than accepting a smaller but surer payoff.

Methods

Participants. We recruited 800 participants from Mechanical Turk, evenly distributed across similarity and knowledge conditions. After eliminating people who failed the comprehension questions, we were left with 550 participants in the final analyses (approximately 46% male, M_{age} = 31.6, SD_{age} = 11.3).

Design and procedure. The design added three components to the $2 condition from Experiment 1.

Similarity manipulation. At the beginning of the experiment, participants answered four questions:

- "Do you prefer more intense kinds of music (e.g., rock or rap) or more mellow kinds of music (e.g., classical or jazz)?"
- "If you had to pick, would you say you are more liberal or more conservative?"
- "How old are you?" [available answers: "I'm 35 years old or older," and "I'm younger than 35 years old"]
- "When making decisions do you tend to rely more on intuition or more on reason?"

In the Similar condition, participants were told that they would be matched with a partner who gave the same answers to three or more of these questions. In the Dissimilar condition, participants are told that they would be matched with a partner who gave the same answers to two or fewer of these questions. Participants were then asked to report how similar they perceived their partner to be to them, on a scale from 0% to 100%.

Big Five personality questionnaire. After participants read the role-playing scenario and made their decision, they were asked to fill out a standard 50-question survey that measured the Big Five personality traits (Goldberg, 1999).

Knowledge-level comprehension questions. Finally, the knowledge-level comprehension questions were administered on a separate page; when answering them, participants were unable to refer back to the initial instructions.

Results and Discussion

Figure 1.5 shows that the Similarity manipulation made no systematic difference. Although ratings of perceived similarity were higher in the Similar condition ($t(548) = 13.87$, $p < .001$), these perceptions of similarity had no effect on the participants' decisions (Wald $\chi^2(1, N = 550) = 0.01$, $p = .944$). We thus collapse across similarity in all other analyses.

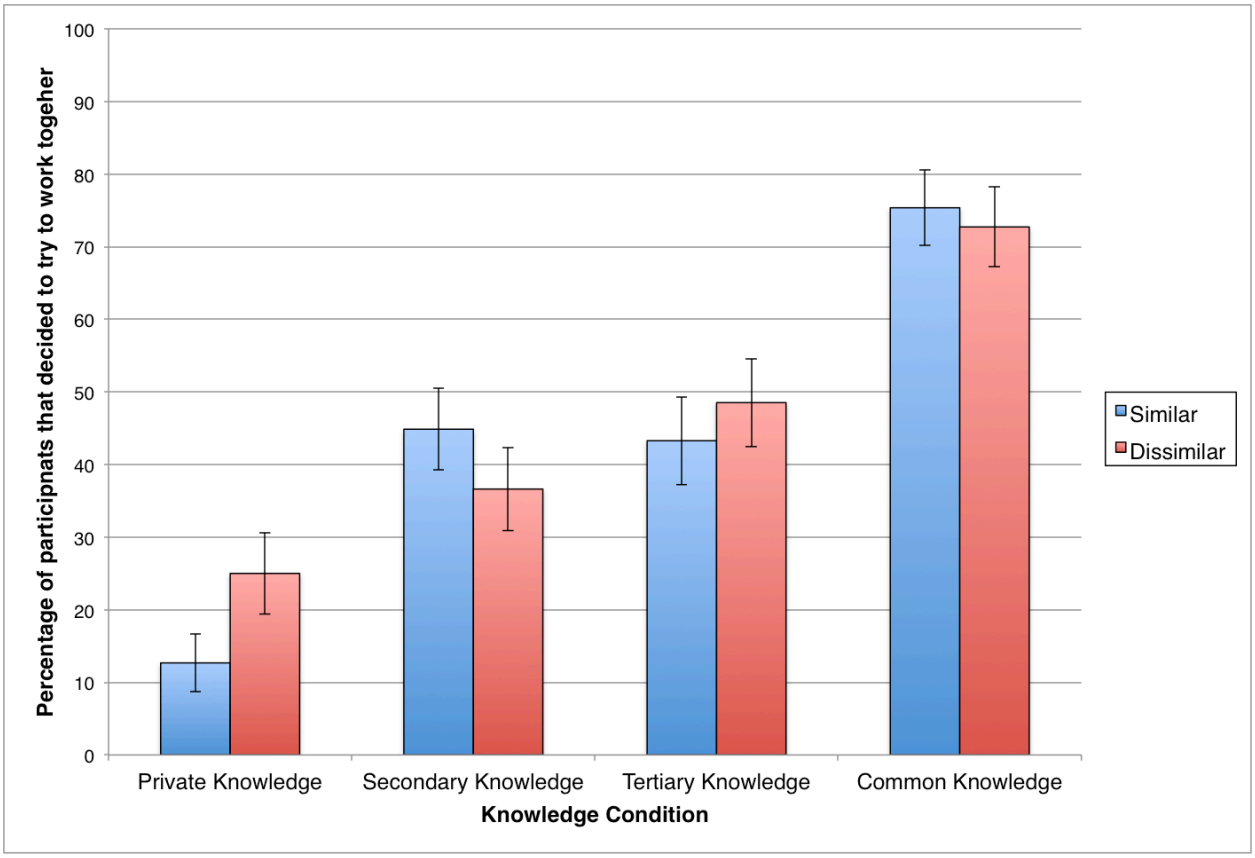

Figure 1.5 Percentage of participants who tried to work together in Experiment 3, organized by knowledge condition and similarity/dissimilarity condition. Error bars represent standard error.

Knowledge level had the same effect as in the first two experiments: More people tried to work together with common knowledge than with tertiary knowledge, $\chi^2(1, N = 270) = 22.28$, $p < .001$, $\phi = .29$, and more tried to work together with secondary than with private knowledge, $\chi^2(1, N = 280) = 16.87$, $p < .001$, $\phi = .25$, but there was no difference between secondary and tertiary knowledge, $\chi^2(1, N = 284) = 0.72$, $p = .397$, $\phi = .05$.

Representations of shared and common knowledge. The confusion matrix for the questions about levels of knowledge is shown in Table 1.3. Participants made significantly more errors with tertiary knowledge than with any of other level of knowledge (planned comparisons: private-tertiary, $\chi^2(1, N = 266) = 33.76$, $p < .001$, $\phi = .36$; secondary-tertiary, $\chi^2(1, N = 284) = 27.91$, $p < .001$, $\phi = .31$; common-tertiary, $\chi^2(1, N = 270) = 13.93$, $p < .001$, $\phi = .23$), and these errors consisted overwhelmingly of misremembering it as secondary knowledge (an error made by 23% of the participants in this condition). Error rates with common knowledge and with secondary knowledge were not significantly different ($\chi^2(1, N = 284) = 2.43$, $p = .119$, $\phi = .09$). None of the other off-diagonal confusions was as high as the one for mistaking tertiary for secondary knowledge. The next highest was 4% (mistaking tertiary knowledge for common knowledge), which was significantly different from the 23% rate for mistaking tertiary for secondary knowledge ($p < .001$).

Table 1.3 Proportion of participants reporting different levels of knowledge in each condition in Experiment 3

Condition	Reported level of knowledge					Proportion
	Private	Secondary	Tertiary	Common	Unclassifiable[a]	Correct
Private	.931	.008	.008	.008	.046	.931
Secondary	.020	.899	.013	.007	.060	.899
Tertiary	0	.230[b]	.637	.044	.089	.637
Common	.007	0	.015	.837	.141	.837

Note. Participants' perceived knowledge level by (actual) knowledge condition. Participants' perceived knowledge level was assessed using comprehension questions. Accurate judgments are those on the diagonal, and are given again in the last column.

[a]Unclassifiable errors correspond to patterns of errors that were logically inconsistent (e.g., reporting that they had tertiary knowledge, but not private knowledge), incomplete, or in which participants reported the correct level of knowledge but chose "can't tell" rather than "yes" for some level of knowledge that they did have.
[b]According to a sign test, tertiary knowledge was mistaken for secondary knowledge more frequently than for common knowledge, $p < .001$.

These results show that higher levels of knowledge are increasingly difficult to represent, as suggested by the theory of mind literature, but only when the knowledge is merely shared; the highest level of all, common knowledge, is almost as easy to represent as the lowest level of shared knowledge. The confusion matrix thus suggests that shared and common knowledge have distinct cognitive representations, but that quantitatively different levels of shared knowledge do not.

Altruistic motives. Figure 1.6 shows that participants in the Shared knowledge conditions who tried to work together scored higher in Agreeableness than those who decided to work alone, a difference not observed in the Private or Common knowledge conditions. Logistic regression, controlling for the main effect of knowledge condition, revealed a significant Knowledge × Agreeableness interaction in coordination attempts, Wald $\chi^2(3, N = 550) = 8.18$, $p = .043$. Post-hoc *t*-tests with Agreeableness as the dependent variable and Decision (work alone vs. together) as an independent variable confirmed that the people who decided to work together with secondary and tertiary knowledge were significantly more agreeable ($t(147) = 2.25$, $p = .013$, and $t(133) = 1.89$, $p = .030$, respectively), but people who decided to work together with private or common knowledge were not ($p > .60$). These results are consistent with the hypothesis that with shared knowledge, people may choose to coordinate with others out of a sense of altruism (perhaps as a signal to encourage coordination in possible future opportunities).

In contrast, Agreeableness did not affect behavior in the common knowledge condition, suggesting that the influence of common knowledge is not driven by reputational concerns.

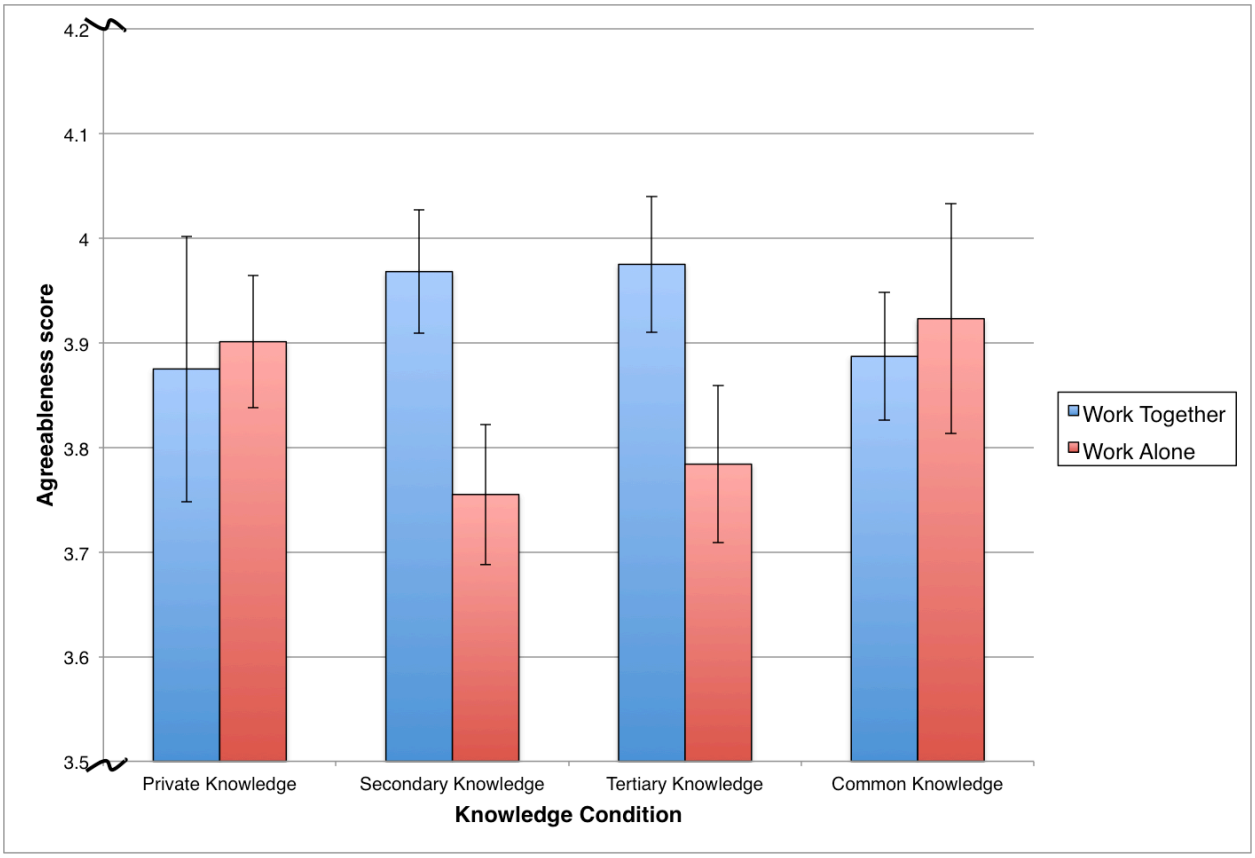

Figure 1.6 Average Agreeableness scale score for participants who tried to work together vs. participants who decided to work alone by knowledge condition in Experiment 3. The figure shows the abbreviated range of 3.5 – 4.2 (full range is 1 – 5). Error bars represent standard error.

Risk-seeking. Figure 1.7 shows a similar pattern for the trait of Openness to Experience (Wald $\chi^2(3, N = 550) = 13.64$, $p = .003$). Coordinators were more Open than non-coordinators when they made their decision with secondary knowledge ($t(147) = 3.30$, $p < .001$), and when they made it with tertiary knowledge ($t(133) = 1.82$, $p = .036$), but not when they made the decision with private or common knowledge ($p > .80$). These results are consistent with the hypothesis that people recognize that, as game theory predicts, attempting to coordinate with

shared knowledge is risky, but attempting to coordinate with common knowledge is not; those who seek risks for gains may thus gamble in conditions of shared knowledge.

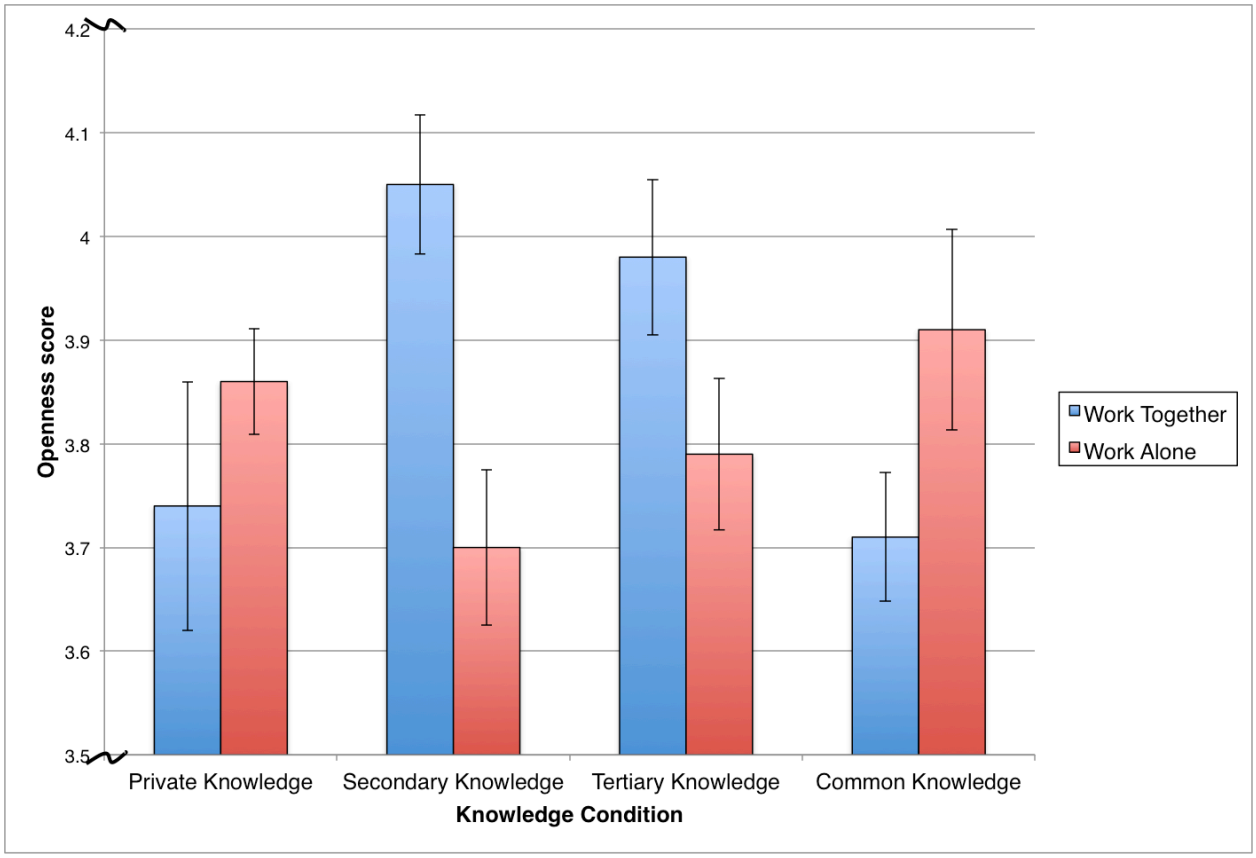

Figure 1.7 Average Openness scale score for participants who tried to work together vs. participants who decided to work alone by knowledge condition in Experiment 3. The figure shows the abbreviated range of 3.5 – 4.2 (full range is 1 – 5). Error bars represent standard error.

No differences were found for the other three personality factors (Extraversion, Conscientiousness, or Neuroticism); all *p*s > .2.

Experiment 4

This experiment was designed to explore how the context in which a coordination problem is framed might affect rates of coordination, and to assess participants' expectations of their partner's decision. Participants engaged in a two-person interaction that was isomorphic to

the $1 payoff condition from Experiment 1, but the scenario literally described a stag hunt involving coordination between hunters.

In this scenario one person was assigned to be an archer, and the other a tracker. They each read that they could always hunt rabbits on their own for a certain payoff of $1, or sometimes hunt deer together to earn $1.10. As in the butcher-baker scenario, they were told they could receive private information from messengers or a public signal, any of which indicated when they could earn more by working together. Replacing the loudspeaker we used in the butcher-baker scenario was a signal fire that could be lit on a hill they could both see and which indicated the presence of deer. Information about participants' level of knowledge was given to them as part of the story rather than in the private and public knowledge boxes that were used in the previous three experiments.

After making their decision, participants were also asked to report what they believed their partner decided to do, in order to test whether participants' decisions were contingent on such guesses, as the logic of a Stag Hunt game suggests is rational. All other aspects of the experiment were the same as in the $1 condition of Experiment 1.

Method

Participants. 400 participants were recruited from Amazon Mechanical Turk (100 per knowledge condition) from the United States to complete a short study for a small payment. After excluding participants that missed comprehension questions about the game's payoff structure, the final sample consisted of 348 participants (40% female, M_{age} = 34.8, SD_{age} = 11.8).

Design and procedure. One participant was assigned to the role of archer, the other to the role of tracker. Participants were told they could either hunt rabbits on their own for a certain payoff of $1 or try to hunt deer together for a payoff of $1.10, which they would only receive if

their partner made the same choice (these are the same payoffs presented in Figure 1.1). In the private, secondary, and tertiary knowledge conditions a messenger delivered the information about the deer being present in the neighboring valley. In the common knowledge condition, participants read that (from the archer's perspective): "You see smoke from the fire pit on the hill that both you and the Tracker can see, which signals to both of you that there are deer in the neighboring valley (so you could potentially earn $1.10 if both you and the Tracker hunt deer). So, the Tracker knows that there are deer today, and knows that you know that there are deer today. The Tracker also knows that you know that he knows that there are deer today, and vice versa."

All subsequent procedures were the same as Experiment 1 (with appropriate rewording), with the addition of a question asking them what they thought their partner decided to do, presented after they made their decision.

Results and Discussion

As in the previous three experiments, more participants decided to try to work together in the common knowledge condition (60%) than in the secondary (33%) or tertiary knowledge conditions (29%) ($\chi^2(1, N = 173) = 12.72, p < .001, \phi = .27$; $\chi^2(1, N = 170) = 16.09, p < .001, \phi = .31$, respectively). Fewer participants decided to try to work together in the private knowledge condition (16%) than in the secondary knowledge condition ($\chi^2(1, N = 178) = 7.35, p = .007, \phi = .20$), and there was no significant difference between the number of participants who tried to work together with secondary knowledge than with tertiary knowledge ($\chi^2(1, N = 173) = .253, p = .615, \phi = .04$).

The logic of a Stag Hunt dictates that one should do what one expects one's partner to do (a rule that we assumed participants obeyed in Experiments 1-3 but without direct evidence).

Indeed, the correlation between participants' decisions and what they reported they thought their partner would do was $r(346) = .80$, $p < .001$. This is perhaps unsurprising, given that a number of previous studies have shown that Mechanical Turk participants believe that they are interacting with a partner as much as participants believe this in the lab, and many classic experiments that critically depend on participant interaction have been replicated on Mechanical Turk (see Amir, et al., 2012; Horton, et al., 2011; Rand, et al., 2012; Summerville & Chartier, 2013; Suri & Watts, 2011).

In sum, the effect of knowledge condition observed in Experiments 1-3 generalized to a different fictional context, and fittingly, the one in which the logic of coordination was historically first elucidated. Moreover, we now have evidence that the key factor predicted to mediate between the knowledge level cued by the environment and people's decision to coordinate—namely the guess that their coordination partner intends to coordinate as well—indeed is critical to that decision.

General Discussion

Humans have lived in large groups throughout their evolutionary history, providing many opportunities for mutually beneficial coordination. Game-theoretic models show that common knowledge has a privileged role in helping individuals solve coordination problems. Taken together, these observations suggest that humans evolved cognitive mechanisms for recognizing common knowledge and distinguishing it from shared knowledge. Our results support this hypothesis. In all four experiments, and with every combination of payoffs, participants were more likely to attempt risky coordination with common knowledge than with shared knowledge. Moreover, coordination attempts with common knowledge did not closely track the cost-benefit

ratio across payoff conditions, indicating that behavior was driven less by estimated probabilities of coordination than by a categorical recognition of a state of common knowledge.

In contrast to the marked distinctions participants made between common, shared, and private knowledge, they made little distinction between different levels of shared knowledge. This is notable because the levels of shared knowledge tested here, secondary and tertiary, span almost the entire range of shared knowledge that people can readily represent, falling just one level short of the four-level maximum observed under favorable conditions by Kinderman et al. (1998). The similar rates of coordination attempts with different levels of shared knowledge is echoed by the pattern of confusions in knowing which state of knowledge was present: People confused tertiary with secondary knowledge, but rarely confused common knowledge with shared knowledge or with private knowledge, nor shared knowledge with private knowledge. This is further reinforced by the finding that people who were higher on the personality traits of Agreeableness or Openness were more likely to attempt risky coordination in both shared knowledge conditions, and only in these shared knowledge conditions.

The cognitive difference between private and shared knowledge has long been established in the literature on theory of mind (Apperly & Butterfill, 2009; Saxe & Young, 2013). The present results suggest that there is also a fundamental difference between shared knowledge and common knowledge. In contrast to their sensitivity to their partner's knowledge state, participants were largely insensitive to the game's payoff structure, responding to changes in expected returns only in an extreme case when successful coordination required three other people to make the same choice.

A puzzle in the results was why a fair number of participants chose to coordinate with shared knowledge. Recall that a participant with secondary knowledge was always matched with

a partner with only private knowledge (they were told that their partner knew the profit for working together, but that their partner was unaware that they also knew this profit). Hence, to coordinate with their partner, a participant should *act in the same way that they would act with private knowledge*. But overall, our participants defied this logic, with more of them opting to work together with secondary knowledge than with private knowledge, providing further support for the claim that participants used the different knowledge states heuristically, rather than rationally calculating their best move.

If they were not playing their best move in cost-benefit terms, why would people risk a certain payout for an uncertain small gain? The personality results indicate that participants may have chosen to try to work together to signal their cooperative character (Agreeableness), or because of a risk-seeking disposition (Openness). Participants did not leverage similarity in making their decisions, but the possibility remains that this manipulation was too cursory to be effective. The rewards for working alone and working together were abstract monetary payoffs, which differ only in magnitude and not in semantic content. However, in many real world situations the nature of the similarity between people is highly relevant to the kind of coordination they are considering, and perceived similarity has been shown to facilitate coordination when coordinating requires matching on some kind of semantic content (Abele et al., 2013). Two roommates, for example, who will have to endure or enjoy each other's music, ought to find similarity in musical tastes more relevant to the decision to live together than similarity in personality or politics.

The pattern of participant payouts provides clues as to why certain aspects of participants' behavior in these experiments were suboptimal. In an arbitrary social situation (such as the artificial scenario of interacting with strangers on the Internet in a contrived game),

it is extraordinarily hard to predict whether coordination will be profitable, because it depends critically on small and unpredictable differences in the decisions of the other participants. In the case of the four-person game above (Experiment 3), coordination turned out to be the *least* profitable strategy, even when a majority of participants chose to coordinate, because even a relatively small number of non-coordinators was enough to scuttle coordination and its rewards. With people unable to predict at exactly which combination of probabilities and payoffs in a given situation this tipping point lies, they may focus predominantly on information that indicates other people's state of knowledge. Actors coordinate when they have evidence for common knowledge, and refrain from coordinating when they do not. For this heuristic to be advantageous in real life, people must have high quality information about common knowledge in ecologically typical environments, and about other people's sensitivity to the same information.

Broader Implications: Common knowledge in Social Life

The finding that people use common knowledge in their decisions to coordinate their behavior, the evidence that common knowledge is a distinct cognitive category, and the suggestion that everyday social life provides reliable cues to common knowledge in opportunities for coordination, all imply that common knowledge has a strong presence in human life and in the phenomena studied by social psychology. This makes it surprising that the psychology of common knowledge has apparently had so little visibility in psychology, and raises the question of whether it is similarly invisible in everyday life, which would be puzzling. If coordination is as important to social life as altruism, and if common knowledge is as indispensable to coordination as reciprocity is to altruism, shouldn't we expect our language and our lives to be permeated with ideas of common knowledge?

We suggest that this is indeed the case, even if it has not been fully appreciated. Just as the logic of reciprocity makes us obsessed with concerns such as debt, favor, bargain, obligation, and so on, we suggest that the logic of common knowledge makes us obsessed with concerns such as publicity, privacy, confidentiality, conventional wisdom, fame, celebrity, hypocrisy, taboo, tact, euphemism, piety, mock outrage, political correctness, and "Washington gaffes" (when a politician says something that is true). In other words, both psychology and everyday social life *have* been concerned with the manifestations of common knowledge, even if psychologists have not hitherto treated them as exemplars of a single principle.

We conclude by suggesting that an acknowledgement of the role of common knowledge in enabling coordination can unify and explain a variety of seemingly unrelated and puzzling phenomena. In particular, much of social life is affected by common-knowledge generators, and much of language and cognition is sensitive to the state of common knowledge.

The most obvious common-knowledge generator is *direct speech.* When one person says something to another in "plain language" or "in so many words," the content of the proposition is common knowledge. Lee and Pinker (2010) showed that when an experimental participant read a vignette in which one person issues an overt threat, bribe, or sexual come-on to another, the participant assumed that each party knows that the other knows that he knows (etc.) the relevant intention, whereas when the same proposition is proffered in an innuendo, even an obvious one, the participant assumes only that the parties know the content of the proposition as private knowledge (e.g., "Michael offered a bribe"), not at higher-order levels (e.g., "Michael knows that the officer knows that he offered a bribe").

The generation of common knowledge may be the function of other deliberate and salient communicative acts. One such example is *joint attention* (Scaife & Bruner, 1975; Tomasello,

1995), in which two people look back and forth at an object and at each other. Joint attention is thought to facilitate the acquisition of words, a classic example of a coordination equilibrium (see Lewis, 1969), and children seem to assume that the meaning of a novel word is common knowledge, but they do not make this assumption about novel facts (Diesendruck & Markson, 2001).

Another class of common-knowledge generators consists of *performatives* (Austin, 1962; Searle, 1989) and the associated phenomenon of *public ceremonies*, in which the public utterance of a proposition (e.g., "I now pronounce you man and wife") ratifies a new coordination equilibrium such as a marriage, law, or court decision. Much of our moral psychology, including moral debate and condemnation, generates common knowledge of prohibited actions, which allows people to coordinate aggression toward wrongdoers (DeScioli & Kurzban, 2013), and may be the logic underlying the well-documented omission bias in moral psychology (DeScioli, Bruening, & Kurzban, 2011). And, as we mentioned in the Introduction, challenges to power that require coordination among many actors are often effected by *public protests*, and increasingly their electronic equivalents.

Common knowledge can also be conveyed nonverbally. Indeed, we propose that the nonverbal signals which accompany self-conscious emotions evolved with their peculiar anatomy and physiological configurations (Tracy & Matsumoto, 2008) precisely because those configurations are simultaneously salient to the expresser and the perceiver (see Provine, 1996, 2012, for discussion). The perceiver knows not only the intended mental state of the expresser but knows that the expresser knows it, that the expresser knows that the perceiver knows it, and so on. Among these nonverbal common-knowledge generators may be the following.

- *Eye contact* is a potent social signal of threats and sexual come-ons precisely because both parties commonly know that they are acknowledging each other's acknowledgment.
- *Blushing* is felt as a somatosensory sensation by the blusher at the same time as it is displayed as a change in skin color to the perceiver. The acute discomfort in blushing resides largely in the knowledge that the blusher knows he or she is blushing, knows that an onlooker knows it, that the onlooker knows that the blusher knows that they know, and so on.
- *Crying* has the same inside-outside salience: A distraught person looking at onlookers through tears cannot avoid the knowledge that others know his tearful state, know that he knows, and so on.
- *Laughter*, with its disruption of the respiration rhythms necessary for speech and its unignorable noise, is also mutually salient to expresser and perceiver.

If this analysis is correct, it predicts that the common knowledge generated by each of these displays is necessary to attain a mutually beneficial equilibrium in a coordination game. (Although such knowledge may be less certain in real-world environments than in the experiments presented here, recall from the Introduction that coordination can be effected by the weaker concept of common p-belief, which generalizes perfectly certain common knowledge to more realistic environments with less certainty.) Pinker (2007), Pinker, Nowak, & Lee (2008), and Lee & Pinker (2010) suggest that one such relevant game is the joint adoption of a Relational Model that consensually governs their interactions, such as communal sharing, authority ranking, equality matching, or market pricing (Fiske 1992, 2004). For example, two people can prosper if they agree to be friends and share things indiscriminately, or if they agree to transact business and one sells something to the other, but not if one believes they are friends

and helps himself to a possession that the other is in the business of selling. In the case of expressions of self-conscious emotions, the game may consist of two parties agreeing that one of them has committed an unintended or regretted harmful act, or is in a vulnerable state, and thus that the second one needn't punish or ostracize him. This equilibrium leaves both of them better off than they would be if the second incurred the cost of punishing or ostracizing the first for a harm he would never repeat anyway (McCullough, 2008). If this theory of nonverbal communication is correct, then expressions that are less likely to generate common knowledge (such as facial expressions which a person can express with little awareness he is expressing it) should not be yoked to an identifiable coordination game.

Since coordination and common knowledge by definition involve multiple parties, we should expect that they manifest themselves not just in small-scale two- or three-person interactions but also within larger groups. That is, the role of common knowledge in solving coordination problems should manifest itself in a number of social-psychological phenomena. Here we can only list the rich possibilities for uniting diverse large-scale societal phenomena as manifestations of coordination problems that involve common knowledge:

- transactive memory (Wegner, 1995; Wegner, Erber, & Raymond, 1991)—which may have interesting parallels with coordinating distributed storage in networked databases (Alberucci & Jäger, 2005; Halpern & Moses, 1990);
- creating and popping market bubbles (Dalkiran et al., 2012; Zuckerman, 2010);
- inaction brought on by diffusion of responsibility (Buchan, Croson, & Dawes, 2002);

- creating and maintaining social norms (Cronk & Leech, 2013), including the sub-optimal norms that result from pluralistic ignorance (Centola, Willer, & Macy, 2005; Willer, Kuwabara, & Macy, 2009);
- the need to define hierarchical roles of leaders and followers when common knowledge is not attainable (Van Vugt, 2006);
- negotiation and bargaining (Ayres & Nalebuff, 1996; Schelling, 1960);
- international relations and diplomacy and so-called "red lines" (Byman, 2013; Hoffman & Yoeli, 2013);
- the subjective perception of currency valuation (Friedman, 1991);
- arbitrary groups, as studied by Social Identity Theorists (Ockenfels & Werner, 2014; Yamagishi, Mifune, Liu, & Pauling, 2008);
- arbitrary rules of etiquette (Schelling, 1960);
- identifiable signals of conspicuous consumption (Veblen, 1899/2007);
- and many other kinds of seemingly arbitrary social constructions (Searle, 1995).

Finally, if common knowledge is a pervasive concern of social life, then it should leave a mark on language in the form of a *conceptual metaphor* (Lakoff & Johnson, 1980; Pinker, 2007): a family of idioms organized around a central image, such as ARGUMENT IS WAR or LOVE IS A JOURNEY. In the case of common knowledge, the central image alludes to the quintessential common-knowledge generator: COMMON KNOWLEDGE IS A CONSPICUOUS OBJECT OR SOUND. Thus, we have a family of expressions which invoke a salient object or event to assert that some proposition or speech act is common knowledge (and hence compels acknowledgment and action by two or more parties), or the converse, that even if some proposition is known by everyone it should strategically be kept out of common knowledge:

The emperor's new clothes.

The elephant in the room.

It's out there; you can't take it back.

It's on the record; to go on record.

The bell can't be unrung (*also*: Some things once said cannot be unsaid).

That's a pretty big matzo ball hanging out there [when one person says "I love you" and the other doesn't reciprocate; from the television show *Seinfeld*]

A bald lie; a barefaced lie [*compare*: a veiled threat; a fig leaf]

To save face; to lose face.

That insult was in his face; he couldn't ignore it.

It's as plain as the nose on your face.

In recent decades, psychologists have recognized that cooperation is one of the hallmarks of the human species, and that its game-theoretic demands have shaped our emotions, our morality, our social relationships, and our language. Much has been learned about these domains of psychology from a focus on the problem of altruistic cooperation and the mechanisms of reciprocity. We hope that comparable insights are waiting to be discovered by psychologists as they investigate the problem of mutualistic cooperation, and as the mechanisms of common knowledge are—as we say—put out there.

PAPER 2

NAVIGATING RELATIONSHIPS WITH LANGUAGE: THE AFFECTIVE AND SOCIAL CONSEQUENCES OF DIRECT AND INDIRECT PROPOSITIONS

Kyle A. Thomas[1], Wendy Berry Mendes[2], Steven Pinker[1]

[1]Department of Psychology, Harvard University

[2]Department of Psychiatry, University of California, San Francisco

Abstract

Transitioning from one type of relationship to another (e.g., from friends to lovers; co-workers to friends) can be awkward if people don't have aligned relationship goals. The recently proposed *strategic speaker theory* (Lee & Pinker, 2010), posits that indirect speech offers an optimal way to make this transition, by allowing speakers to assess another person's interests, while minimizing damage to an existing relationship. This experiment tests two predictions of the theory: (1) Affective reactions track the social costs and benefits of indirect and direct propositions, and (2) Indirect speech helps prevent rumors from spreading to third parties. Pairs of recently acquainted male and female participants acted out two scenarios, one involving a sexual proposition, the other a bribe proposition. The propositions were made either directly or indirectly, while participants' cardiovascular responses were measured to assess affective responses. Results largely support the predictions. Participants' responses to direct and indirect propositions were different in both scenarios, and were moderated by variables that uniquely affect the costs and benefits of the two scenarios (gender in the sexual proposition scenario; proposer or receiver in the bribe scenario). Participants' self-reported responses also showed that while they understood the intended meaning of the indirect sexual proposition, they would not pass it on unambiguously to a friend. These findings provide the first evidence that affective reactions track the costs and benefits of direct and indirect speech, and that indirect speech helps prevent rumors.

Introduction

When someone has a crush on a friend, but isn't sure if their friend feels the same way, they frequently receive the advice to just tell them. Yet, this advice is ignored almost as frequently as it is given, because baldly stating one's feelings can be terrifying, and people rightly worry that doing so could destroy their friendship. Instead, people generally try to figure out a way to communicate their desire indirectly, and even using a completely transparent innuendo can feel much less threatening. Furthermore, if a veiled offer to take the relationship to the next level is rejected, the two may remain friends the next day, even if the intended message couldn't possibly have been misunderstood, yet the friendship may be unsalvageable if the offer was made directly.

Why can it be so terrifying to state one's desires directly? How does indirect speech make this less scary? And, why might a direct proposition destroy a friendship, when a completely transparent innuendo would not? While we all share the intuition that indirect propositions often feel much safer than direct ones, it is not immediately clear why this should be the case. A common folk explanation focuses on the fact that indirect propositions do not create 100% certainty, yet this merely begs the question, because it fails to explain how an indirect proposition that creates 99% certainty can lead to dramatically different outcomes than a direct one. Why don't the consequences scale with the level of certainty? Why is there such a dramatic discontinuity between indirect and direct propositions?

Theoretical Background: Indirect Speech, Cooperation, and Politeness

In everyday conversation, people speak indirectly all the time, a tendency that is so natural and intuitive that it often goes unnoticed until one is forced to be direct, precise, and unambiguous, such as in legal settings, or when programming a computer. At least since Grice

(1975), explanations for how people understand indirect speech generally start with the assumption that conversational partners cooperate to achieve a shared goal of communication (see also, Clark, 1996). From this *cooperative principle*, Grice derived four *conversational maxims*: Listeners assume that speakers should be (1) succinct, (2) truthful, (3) relevant, and (4) clear. Grice argued that these assumptions help listeners infer the meaning of an indirect utterance, by providing them with a guide for how to fill in any missing information when a speaker violates one of the maxims.

While Grice explained how listeners might understand indirect speech, he offered no explanation for why people might speak indirectly in the first place. Politeness theory, proposed by Brown & Levinson (1987), aimed to fill this gap, by positing that when imposing on others, people violate the maxims to add layers of politeness to their communicative acts. Politeness theory integrates Grice's insights with those of Erving Goffman (1955, 1959), who argued that much of human social behavior can be understood as efforts to establish and maintain *face* (i.e. prestige and autonomy, from the idiom *to save face*). Brown and Levinson argued that people often use indirect speech to help both conversational partners maintain face, such as when a speaker makes a veiled request (e.g., "It would be awesome if you were around to help me move this weekend") rather than stating their request directly (e.g., "Please help me move this weekend").

Research has borne out many of the predictions of politeness theory, but there are both empirical and theoretical reasons to doubt that politeness is the only reason people use indirect speech to navigate their relationships. Empirically, some of the predictions of politeness theory have been contradicted (Holtgraves & Yang, 1990; Lee & Pinker, 2010). For example, no one would consider a veiled threat polite (e.g., "Nice shop you've got here; sure would be a shame if

something happened to it"), nor would anyone interpret a passive aggressive jab as an effort to save the listener's face (e.g., "I guess I'll just clean the bathroom again then…like always"). Theoretically, it is dubious that conversational partners are always engaged in a purely cooperative relationship, as biological signaling theory suggests that communicating organisms often have conflicting interests (Maynard-Smith & Harper, 2003; Dawkins & Krebs, 1978). In the examples given above, the speaker's interests clearly diverge from those of the listener.

Indeed, the kinds of indirect speech acts that don't conform to the predictions of politeness theory generally occur in situations where conversational partners have potential conflicts of interest (Lee & Pinker, 2010). A speaker who delivers a veiled bribe for example, is often uncertain whether the other person shares their interest in making such an illicit transaction. The same is true when one conveys a sexual proposition through innuendo: The person being propositioned may not share the speaker's desire for a sexual relationship.

The Strategic Speaker Theory: Managing Uncertainty and Conflicts of Interest

Both theory and research suggest that even in the most cooperative relationships, people's interests are never perfectly aligned (Richardson, 2014; Trivers, 1985). For example, married couples often share the cooperative goal of raising their children, but they may frequently disagree on how much each should contribute towards this goal. People establish and maintain relationships by identifying and focusing on shared interests, while avoiding or managing any conflicts of interest, which can destroy a relationship if they become insurmountable (Argyle & Furnham, 1983; Helgeson, Shaver, & Dyer, 1987; McPherson, Smith-Lovin, & Cook, 2001; Walster, Berscheid, & Walster, 1973). This task is complicated by the fact that other people's mental states—their interests, motivations, goals, beliefs, etc.—are not directly observable, which can make it difficult to assess how other's interests overlap or conflict

with one's own (Baron-Cohen, 1997; Gilbert, 1998; Whiten & Byrne, 1997). One way to assess someone else's interests is by talking with them; however, this may expose a previously unrecognized conflict of interest. To understand how this might pose a dilemma, imagine two flirtatious friends who have never been romantically involved. If one person wants to shift from being friends to lovers, they face a catch-22: They may have to reveal their desire to assess how their friend feels, but doing so risks revealing a conflict of interest if their friend doesn't feel the same way, which could potentially ruin the friendship.

The recently proposed *strategic speaker theory* (hereafter, SST), posits that managing these two concerns depends on different kinds of knowledge, and that by using indirect speech, speakers can exploit this difference to assess the interests of their conversational partner, while minimizing damage to any existing relationship (Pinker, 2007; Pinker, Nowak & Lee, 2008; Lee & Pinker, 2010). Assessing another person's interests only requires *private knowledge*, as demonstrated by the fact that this can easily be done without the other person knowing, such as by surreptitiously overhearing them talking with someone else. In contrast, the theory posits that damaging a relationship often requires *common knowledge*, which is defined as an infinite recursion of shared beliefs, such that A knows X, A knows that B knows X, A knows that B knows that A knows X, and knows that B knows that they know that B knows X, ad infinitum. Previous research has confirmed that indirect speech can create private knowledge of a proposition without making it common knowledge, unlike direct speech, which makes a proposition both privately and commonly known (Lee & Pinker, 2010). To understand how this property of indirect speech might help a speaker navigate their dual concerns requires a brief delve into the game theory of *coordination games*.

Common Knowledge and Coordination

In a coordination game, two or more people can benefit by coordinating their behavior, but there is more than one way to do so, and everyone incurs a cost if they fail to coordinate. For example, everyone benefits by driving on the same side of the road, but this can be either the left side or the right side; likewise, dissidents can benefit through collective action, but their goals may be pursued in a number of different ways. In both cases, there can be steep penalties for coordination failure—a car accident in the first case, persecution or wasted effort in the second.

When trying to coordinate with someone else, an individual should do whatever his or her partner does, which can pose a problem, because each person's best option depends on what they expect the other person to do. Recognizing this mutual dependency, an individual should only expect their partner to choose some option if they think their partner expects them to, and only if they expect their partner to expect them to expect this, and so on. This infinite recursion of expectations is the reason why an infinite recursion of shared beliefs (i.e., common knowledge) is often needed to coordinate (Chwe, 2001; Lewis, 1969; Schelling, 1960). Recent research has shown that people are at least implicitly cognizant of this requirement, and that they represent common knowledge as distinct from lower, finite levels of shared beliefs (Thomas et al., 2014).

Game theorists have demonstrated that common knowledge is both necessary and sufficient to shift from one coordination option to another (Dalkiran et al., 2012; Morris, 2002; Rubinstein, 1989)[6]. For example, if two friends meet up at the same time and place every week, they can only switch locations if they establish common knowledge that they will do so, because

[6] Technically, switching between coordination options can be achieved with the weaker state of *common p-belief*, in which individuals' knowledge need not be 100% certain (Dalkiran et al., 2012; Monderer & Samet, 1989). However, the same dynamics apply, and here we use the term *common knowledge* broadly, to also include "sufficiently high common p-belief".

neither one wants to end up at the new location while the other is at the old location. More importantly for indirect speech, making a conflict of interest common knowledge can lead to coordination failure, or force people to coordinate on some option that none of them wants (Dalkirin et al., 2012). Continuing with our example, two friends may not be able to meet up if it becomes common knowledge that one person would like to change the time or place while the other wouldn't, and they fail to establish a final commonly-known agreement on one of the two options.

To understand this dynamic, imagine two friends, Emily and Kate, who get together every Wednesday at a local café, and one Wednesday Emily ends up on the other side of town shortly before their regularly scheduled time. Emily calls to see if Kate can meet at a café closer to her instead, but Kate says she'd rather not have to drive across town. Then, before they can resolve the issue, Emily's cell phone dies. Because they both expressed a different preference, but failed to come to an agreement on where to meet before the phone died, neither can be sure where they should go. Furthermore, if the worst outcome would be sitting at a café alone, both may decide to just scrap this week's meet-up. By exposing their conflict of interest over where to meet, Emily disrupted their well-coordinated tradition, and because they failed to reestablish common knowledge of where to meet, they may both abandon this week's meeting entirely rather than risk wasting their time. Although both might prefer meeting at either location to not meeting at all, each may abandon the meeting simply because they expect the other one to do so, or expect the other one to expect them to do so, and so on.

Relationships as Coordination Games

The SST models relationships as a kind of coordination game because people can have many different types of relationships (e.g., friends, lovers, co-workers, or acquaintances), and

each relationship type is governed by distinct expectations and norms (Fiske, 1992; Haslam, 2004). While two people can often benefit by coordinating on one type of relationship or another, if they can't agree on what the prevailing relationship should be, their best option may be to have no relationship at all. For example, an individual who commits a lot of time, energy, and resources in trying to establish a romantic relationship with a friend who isn't interested, may be better off abandoning the relationship entirely. Conversely, the benefits of maintaining a friendship with someone whose love is unrequited may not be worth the constant harassment and annoyance that often typify such asymmetrical relationships. Furthermore, if someone else desires to opt out of having any kind of relationship, typically one's best option is to do the same, because investing in a one-sided relationship is worse than abandoning it altogether.

Because each person is better off abandoning a relationship if they expect the other person to, mutually opting out of any relationship is also a kind of coordination. Just as Emily and Kate had to decide not only between going to the normal café or the new café, but also abandoning the effort entirely, people often must decide not only among the various types of relationships they could engage in, but also have the option to avoid pursuing any kind of relationship at all. And, since coordination depends on mutual expectations, this can lead to a paradoxical situation in which two people abandon their relationship simply because each one expects the other one to, even if neither wants to. This dynamic is perhaps most obvious in teenage dating, when for example, a girl dumps her boyfriend, not because she wants to, but because she thinks her boyfriend is about to break up with her. Similarly, two people who both want to remain friends may end up failing to do so, simply because each expects the other to jump ship, and neither wants to be the last one to do so.

The Strategy of Indirect Speech

If relationship negotiation follows this logic of coordination, indirect speech could provide speakers with an optimal strategy to simultaneously assess whether their conversational partner shares their interests, while hedging against ruining the relationship if they do not. If two friends go out one night, and one invites the other up for coffee at 2am, this creates an opportunity to transition from friends to lovers if the other person is interested, but can also allow them to remain friends the next day if the offer is rejected. In either case, this interaction gives both people private knowledge of the other's interests—people rarely drink coffee at 2am—but it does not make a conflict of interest common knowledge if the offer is rejected (Lee & Pinker, 2010). If instead the speaker made the proposition directly (e.g., "Would you like to come up and have sex?"), and it was rejected, this would make it commonly known between both people that one wants to be lovers while the other does not. By making this common knowledge, a direct proposition may make the friendship unsalvageable, even if both still wish to remain friends, since either or both may bail on the friendship solely because they expect the other person to, or think that the other person expects them to, and so on. By preventing common knowledge of disparate interests, an indirect proposition can allow the two to remain friends if they both still wish to.

These insights are based on formal mathematical models, in which abstract rational agents calculate their best move based on explicit numerical payoffs, but the models are agnostic as to how these payoffs are instantiated. Translating the models to real human behavior thus requires some psychological mechanism that captures the payoffs in some way. People's social behavior is generally driven not by conscious calculations of precisely enumerated payoffs, but rather by emotions and motivations that evolved to track the perceived costs and benefits of different outcomes (albeit, imperfectly). We establish beneficial relationships because this makes

us feel good, and avoid damaging valuable existing relationships because doing so makes us feel bad (Baumeister & Leary, 1995). The more valuable a relationship is, or can be, the better we feel creating it and the worse we feel if it is destroyed. The SST hypothesizes that indirect and direct propositions are associated with different affective states that track the costs and benefits of different potential outcomes.

Indirect Speech, Third Parties, and Rumors

The SST further proposes that indirect speech can help speakers prevent rumors from spreading to third parties about their interests and intentions. Indirect propositions often require context and non-linguistic cues to interpret the intended message. An invitation for coffee at 2am generally conveys something very different than an invitation for coffee mid-afternoon, and an otherwise innocent request can take on an entirely different meaning when combined with the right tone of voice or body language (a phenomenon often exploited by comedians). Just as retelling a hilarious joke verbatim may not have the same effect without the appropriate delivery, it can be difficult to share what was conveyed through an indirect proposition if one cannot recreate the original context in which it occurred.

The SST suggests that this is because the discrete nature of words, phonemes, and syntax make direct speech a *digital* medium of communication, while many of the contextual and non-linguistic cues needed to interpret indirect propositions are *analog* (Lee & Pinker, 2010; Pinker, Nowak, & Lee, 2008). Digital information can generally be reproduced with much higher fidelity than analog information because small variations in signal quality can be corrected at each reproduction, slowing the inevitable creep of entropy in the signal each time it is reproduced (Tocci, Widmer, & Moss, 2007). This is why music that is stored digitally (e.g., on a CD or as an mp3) can be copied indefinitely with minimal loss of quality, but serial copies degrade rapidly if

the music is stored in an analog format (e.g. a cassette tape or record). Similarly, the analog features required to interpret indirect propositions may make them difficult to pass on to third parties, or for third parties to pass on to fourth parties, and so on. As Pinker, Nowak, and Lee (2008) state, "Overt propositions, in contrast [to indirect propositions], may be perceived as *context-free*—their intent can be interpreted by eavesdroppers—and *lossless*—their intent can be transmitted perfectly along chains of gossipers (in the same way that other digital media, such as files of music or images, can be transmitted losslessly)" (p. 837, italics in the original).

Strategic Indirect Speech is Difficult to Study in the Lab

The very nature of strategic indirect speech makes it difficult to study in the laboratory, because risky propositions are typically delivered in private or secluded settings, and people are generally loath to make them directly. Indeed, we have previously struggled to elicit direct propositions from participants in the lab (see Footnote 6 in Lee & Pinker, 2010). Moreover, measuring affective responses to acts that have plausible deniability, are embarrassing, or may tarnish positive self-representations adds further obstacles to understanding how people respond to using and receiving propositions that vary in directness. To bring the phenomenon into the lab, we had participants act out two scripted role-playing scenarios involving common situations that could plausibly elicit a risky proposition (one involved a sexual proposition, the other a bribe proposition). To measure affective responses to indirect and direct propositions, we employed psychophysiological methods to capture online changes in cardiovascular reactivity. These measures allow for continuous, dynamic, and online assessment of affective states, without interrupting the interaction, or relying on self-reports.

Specifically, we employed a set of cardiovascular measures that have been consistently linked to the well-established affective states of *challenge* and *threat* (Blascovich & Mendes,

2010; Mendes & Park, 2014). Both of these affective states occur in the context of motivated performance situations—defined as self- or goal-relevant, active tasks that require instrumental cognitive responses—but are differentiated by how people assess and react to the situation, and have distinct profiles of cardiovascular responses. Challenge states can be conceived of as "good stress", as they occur when individuals appraise their available resources to exceed the demands of the task, and they are characterized by an approach motivation, typically better cognitive performance, and greater efficiency in physiological mobilization (i.e., increased blood flow to the periphery and brain). Threat states can be conceived of as "bad stress", as they occur when individuals perceive the demands of the task to exceed their available resources, and they are characterized by an avoidance motivation, poorer cognitive performance (though better danger detection), and more maladaptive physiological responses (i.e., increased vasoconstriction and less overall blood flow). Both states involve an increase in sympathetic nervous system activity, which is indicated by an increase in *ventricle contractility* (VC, the force of the left ventricle contraction). However, challenge states are characterized by an increase in *cardiac output* (CO, amount of blood pumped out by the heart per minute), and a decrease in *total peripheral resistance* (TPR, a measure of overall vasodilation), whereas threat states are characterized by little or no increase in CO, and an increase in TPR (Mendes, 2009; Mendes & Park, 2014).

The challenge and threat framework is useful for our purposes here, because these cardiovascular reactivity profiles are well established, and the framework can be recast in terms of perceived costs and benefits to map onto the SST framework[7]. If an individual faces an

[7] While there is a strong correspondence between affective states and perceived costs and benefits, we note that this mapping is imperfect for a few reasons. Intangible costs and benefits can never be assessed perfectly or precisely, may occur below conscious awareness, and other psychological factors may impact how these assessments are

interpersonal task they don't think they can manage—the prototypical situation in which threat states manifest—the perceived costs of engaging in the task outweigh the benefits. In contrast, when an individual faces an interpersonal task they do think they can manage—the prototypical situation in which challenge states manifest—the perceived benefits outweigh the costs. Thus, in the context of this experiment, threat states can be interpreted as indirect evidence that participants perceived the costs of a proposition to outweigh its benefits, and challenge states as indirect evidence that participants perceived the benefits to outweigh the costs.

The Present Research

In this experiment, pairs of male and female participants interacted in two role-playing scenarios, in which they both read lines from a script, while we measured their physiological reactions to different kinds of propositions. In the *date scenario*, participants acted out a date, which ends at the front door of one of their houses. One participant was randomly assigned to make a direct sexual proposition, make an indirect sexual proposition, or ask whether their partner would like to go out again sometime. In the *restaurant scenario*, one participant played the role of a patron trying to get a table at a fully booked restaurant, and the other participant played the role of the maitre d'. After a conversation in which the maitre d' informed the patron that there were no available tables, the patron was randomly assigned to make a direct bribe proposition to get a table, make an indirect bribe proposition, or ask if the maitre d' could recommend another restaurant. These scenarios were chosen because they provide familiar situations in which people commonly make such risky propositions, and often make them indirectly (see, Feiler, 2000, for a fanciful recounting of trying to bribe maitre d's).

Affective Reactions

---

translated into affect, such as personality variables, past experience, or residual or incidental affect (see Blascovich, Mendes, Tomaka, Salomon, & Seery, 2003).

The SST proposes that indirect speech offers speakers an opportunity to shift to a new desired relationship, while avoiding potential damage to an existing relationship, and that affective reactions to different kinds of propositions should track perceived costs and benefits. Combining this with the challenge and threat framework predicts that participants should show threat responses when they perceive the costs of a proposition to outweigh its potential benefits, and show challenge responses when they perceive the benefits of a proposition to outweigh its costs. We thus hypothesized that direct propositions should generally lead to threat responses (increases in VC and TPR, with no increase in CO), while indirect propositions should generally lead to challenge responses (increases in VC and CO, and decreases in TPR). We note however, that the theory predicts that direct propositions may sometimes be preferred—and thus lead to a challenge state—when the potential costs of making a direct proposition are lower than the potential costs of being misunderstood (e.g., if all cops are known to be corrupt, offering a bribe directly is better than risking a possible misunderstanding; see Figure 1 in Lee & Pinker, 2010). This general hypothesis thus assumes that participants perceive the costs of direct propositions to outweigh their benefits in the scenarios they acted out. Adding further complexity to this general prediction, we also predicted that participants' physiological reactions would track contextual and demographic variables that uniquely affect the relative costs and benefits in the two scenarios.

Insights from evolutionary biology and research on human sexuality have shown how the potential benefits and costs from a casual sexual encounter tend to be different for males and females, at least on average (Buss, 1994; Clark & Hatfield, 1989; Li & Kenrick, 2006; Rusbult, Martz, & Agnew, 1998; Symons, 1979). Males are frequently highly motivated to seek out casual sexual encounters, which generally present them with a large potential reproductive

benefit that outweighs potential costs. However, the cost-benefit calculus for females is much more variable and complex, with the costs typically outweighing the benefits, in terms of reproductive output, their motivation to pursue them, and the potential reputational costs they might incur. This asymmetry predicts that males and females will exhibit different responses to the sexual propositions. Specifically, females will exhibit stronger threat responses than males to the direct proposition, while males will be more likely to exhibit challenge responses, especially to the indirect proposition.

In contrast, when offering a bribe to a maitre d', the costs and benefits are generally the same for males and females. However, the costs and benefits are different for a patron who offers a bribe and a maitre d' who receives the proposition, especially if the maitre d' does not accept the bribe (as was the case here—the bribe was neither accepted nor rejected). This asymmetry predicts that participants should respond differently to the bribe proposition when they are the patron, than when they are the maitre d'. We thus predicted that patrons would have stronger threat responses than the maitre d' in the direct proposition condition.

To summarize, the SST predicts that participants will exhibit different affective responses to direct and indirect propositions, and that these responses will track the perceived costs and benefits, such that direct propositions should generally be more threatening than indirect ones. Sex should moderate these effects in the date scenario, but not the restaurant scenario, such that females will have stronger threat responses to the direct sexual proposition, while males will be more likely to have challenge responses, especially to the indirect proposition. Whether a participant makes the proposition or receives it should moderate their physiological responses in the restaurant scenario, such that participants who make the direct bribe proposition (patrons) will show stronger threat responses than participants who receive it (maitre d's).

Preventing Rumors

To test whether indirect speech also helps speakers prevent rumors from spreading to third parties, participants answered a series of questions about the date scenario to assess their perceptions of what was asked, what the speaker intended, and what they would report about the date to a friend. The SST theory predicts that while participants would understand the intended proposition regardless of whether it was made directly or indirectly, they would only report it directly and unambiguously to their friend when it was made directly.

Method

Participants

One hundred and eleven dyads ($N = 222$), each consisting of a male and a female, were recruited from Harvard University's subject pool ($n = 144$) or the San Francisco Bay Area ($n = 68$). The majority of participants were undergraduates at Harvard or other institutions in Boston or San Francisco ($M_{age} = 22.2$, $SD_{age} = 3.7$). Participants received either $25, or course credit for their participation. Ten participants were excluded a priori from analyses because of problems noted at the time of data collection, such as florid psychopathological symptoms or experimenter error, yielding a final sample of $N = 212$.

Procedure

Participants were brought into separate rooms, where they gave informed consent, completed a brief questionnaire on their subjective social status (see Adler, Epel, Castellazzo & Ickovics, 2000) and current affect (assessed using the PANAS, see Watson, Clark & Tellegen, 1988), gave a saliva sample, had sensors attached to measure their cardiovascular responses, and then sat and relaxed for five minutes while baseline cardiovascular measures were recorded.

They were then brought into a room together, given a set of questions to discuss for five minutes to get to know each other[8], and then sat in silence for a two-minute resting period.

Participants then engaged in two back-to-back experimental interactions. In each of these seven-minute interactions, participants first role-played a scripted scenario for about three and a half minutes, and then played a cooperative game (based on the board game Taboo) for the remaining time. The game provides a cooperative and active task that allowed us to continue measuring participants' cardiovascular responses evoked by the role-playing scenarios (see Mendes, et al., 2007; Mendes et al., 2008).

In order to maintain experimental control over the timing of participants' reactions to events in the role-playing scenarios, the scripts were read from booklets with each line on a separate page, so participants couldn't read ahead. After each of these seven-minute interactions, participants sat in silence for a two-minute resting period.

Each participant dyad was assigned to one of three speech conditions in the scenarios, involving either: (1) a direct proposition, (2) an indirect proposition, or (3) a neutral control question. In the date scenario, participants acted out a dinner date that was going well, which then ended at the front door of one of their houses, providing a natural context for a sexual proposition. In the direct condition, one participant then asked, "Would you like to come up and sleep with me?" In the indirect condition, they asked, "Would you like to come up for a drink?" And, in the neutral control condition, they asked, "Would you like to go out again next Thursday?" In the restaurant scenario, one participant played the role of a patron trying to get a table at a fully booked restaurant, while the other played the role of the maitre d', a context in

[8] These were drawn from the Fast Friends Procedure, and included questions such as, "Would you like to be famous? In what way?" (see, http://www.ocf.berkeley.edu/~rascl/tools/secondFastFriendsQuestions.html)

which a patron could plausibly offer the maitre d' a bribe to be seated right away. In the direct condition the patron then flashed a $50 bill that was supplied in between the pages of the script in order to make the propositions coherent, and asked, "If I give you $50 will you seat us right away?" In the indirect condition the patron flashed the $50 bill and asked, "Do you think there might be a cancellation?" And, in the neutral condition, the patron asked, "Could you recommend another restaurant?"

After the indirect and direct propositions in both scenarios, the scripts included a few additional lines that provided a tacit acknowledgment of the proposition that was just made (e.g., "uuuum…I have a long day tomorrow…" after the sexual proposition; "I'm not really allowed to do that…" after the bribe proposition), but left it unresolved, so that the proposition was neither accepted nor rejected. These lines were identical in the indirect and direct conditions; however, they were different in the neutral conditions, which was necessary to make the dialogue coherent (e.g., it wouldn't make sense for a maitre d' to discuss a bribe after someone asks them to recommend another restaurant).

The sex of the participant who made the proposition, the speech condition for each scenario, and the scenario order were all randomly assigned in a 2 × 3 × 2 mixed design with speech condition as a within-subjects variable, which was counter-balanced.

Participants were then brought back into separate rooms, all sensors were removed, and they completed another questionnaire. This questionnaire again asked them to report their social status and current affect, as well as whether either of the scenarios made them uncomfortable (to assess whether they perceived the costs to outweigh the benefits in either scenario). Participants were then asked to report what was asked at the end of the date (a manipulation check), and what they thought the propositioner would have wanted and what would have likely happened if they

agreed to the request (used to assess how they interpreted the intended message). Then all participants were asked to imagine that, over lunch the next day, their best friend inquires how their date went, and to write down what they would tell their friend in a brief paragraph. These items focused solely on the date scenario because sexual propositions, interests, and behavior are very common topics of gossip, unlike bribes.

After participants completed the questionnaire, they were paid and debriefed.

Cardiovascular Measures

Cardiovascular measures were recorded noninvasively according to established guidelines (Sherwood et al., 1990). We measured electrocardiography (ECG100C) and impedance cardiography (HIC 2000) and integrated the signals with an MP150 system (BIOPAC Systems Inc, Goleta, CA). Sensors to measure ECG were applied in a standard lead II configuration and impedance cardiography was obtained using four mylar bands that completely encircled the neck and chest area. A 1 mA AC current at 100kHz was passed through the outer bands, and Z0 and dZ/dt were recorded from the inner bands. All signals were filtered online and collected at 1000Hz.

We used IMP (3.0) modules from Mindware Technologies (Gahanna, OH) to edit and score the physiological parameters of interest, specifically cardiac output (CO) and pre-ejection period (PEP). PEP, a chronotropic measure of the strength of the contractile force of the heart, provides a sensitive index of sympathetic control of the heart (Brownley, Hurwitz, & Schneiderman, 2000). PEP values were multiplied by -1 to yield ventricle contractility (VC). Impedance data were ensemble averaged with 1-minute epochs and each waveform was verified and edited as needed. Visual inspection of the waveforms focused on detecting ectopic beats and accurate Q, R, S placement in the ECG trace, and accurate detection of the B-, X-, and Z-points

(aortic valve opening, aortic valve closing, and dZ/dt max, respectively) on the dZ/dt waveform. We did not rely on available B-detection algorithms (e.g., Lozano et al., 2007), but rather only accepted waveforms in which we could visually confirm a B-point.

We also attempted to collect blood pressure responses given the previous data on TPR differences between challenge and threat states. Unfortunately, the diastolic blood pressure (DBP) data obtained was unreliable, as there was large variability across consecutive time points, and many invalid values that were physiological implausible. These problems with obtaining DBP responses during the social interaction were likely due to excessive movement and fidgeting during the script reading; when participants move their arms or wrists, blood pressure responses can produce highly variable estimates of questionable validity (see Berntson, Quigley, & Lozano, 2007; Mendes, 2009). Because of these problems we could not calculate TPR, which requires valid measurements of both systolic blood pressure (SBP) and DBP (see, Mendes, 2009). However, we were able to salvage SBP data, which was cleaner than the DBP data.

Based on our hypotheses and the valid data collected, we focused on reactivity (i.e., changes from baseline to stress task) of three autonomic nervous system parameters: CO, VC, and SBP. We calculated reactivity scores by subtracting the last minute of the baseline period from the critical minute of the task period—the minute just after the proposition was delivered by the proposer. We used the last minute of the baseline period because that is typically the time when autonomic parameters are at their most quiescent state and the subsequent minute following the proposition as the proximal response window once the proposition was made (see Mendes, 2009 for details).

Results

Self-report Data

Positive and negative affect. Two positive affect scale scores and two negative affect scale scores were calculated for every participant, based on their responses to the two PANAS questionnaires given before and after the experimental interactions (see Watson, Clark & Tellegen, 1988). Difference scores were then calculated by subtracting each participant's pre-interaction scale scores from their corresponding post-interaction scale scores. Two 3 × 2 × 2 ANOVAs (Speech Condition × Sex × Role) were run to analyze whether any of these variables from either scenario affected the two difference scores. No significant results were found for the date scenario, with all *p*s > .089, and all *p*s > .262 for any term in the equation that included the critical variable of speech condition; nor were any significant results found for the restaurant scenario, with all *p*s > .138, and all *p*s > .375 for any term that included speech condition.

Discomfort in the two scenarios. In reply to a free-response question that asked whether either of the scenarios made them uncomfortable, 21% of participants reported that the date scenario did, while only 3% said the restaurant scenario did, a difference that a Wilcoxon Signed-ranks test showed was highly significant, $Z = 5.11$, $p < .001$, $r = .35$. Furthermore, while the restaurant scenario was reported as uncomfortable by an equivalently low percentage of participants in the neutral (3%), indirect (3%), and direct (4%) conditions, $\chi^2(2, N = 207^9) = .31$, $p = .861$, $\phi = .04$, more participants reported that the date made them uncomfortable when they were in the direct condition (27%), than in the indirect (21%) or neutral conditions (15%). Although this difference was not statistically significant, $\chi^2(2, N = 209) = 3.30$, $p = .192$, $\phi = .13$, this statistic collapses across males, who showed no significant difference across conditions, $\chi^2(2, N = 102) = 2.27$, $p = .322$, $\phi = .15$, and females, who did, $\chi^2(2, N = 107) = 6.05$, $p = .049$, ϕ

[9] While all participants engaged in both the date and restaurant scenarios, and received the same questionnaire, the numbers of participants reported in the analyses in this section are not perfectly aligned with recruitment numbers because of a few missing data points.

= .24. A higher percentage of females reported that they were uncomfortable with the direct sexual proposition (41%)[10], than both the indirect (19%), $\chi^2(1, N = 75) = 4.43$, $p = .035$, $\phi = .24$, and neutral propositions (18%), $\chi^2(1, N = 69) = 3.85$, $p = .050$, $\phi = .24$. Furthermore, a higher percentage of females (41%) than males (14%) reported that the direct sexual proposition made them uncomfortable, $\chi^2(1, N = 74) = 6.85$, $p = .009$, $\phi = .30$, which was not the case in either the neutral or indirect conditions (both *p*s > .40). To summarize, a higher percentage of females reported being uncomfortable with the direct sexual proposition than any group from any speech condition of either scenario.

Perceptions of the date propositions and spreading rumors. An experimenter and another outside rater, both blind to condition, coded participants' responses to the three free-response questions about the date proposition, to assess whether participants made any reference to sex, and if so, whether this reference was direct (e.g., "he asked me to sleep with him"), or indirect (e.g., "he invited me up to his place"). The criterion used to determine whether a reference was direct or indirect was whether or not any obvious implied reference to sex could plausibly be *canceled* by some follow-up statement (e.g., "he invited me up to his place" can be canceled by "…but only to help him get his door unlocked"; whereas, "he asked me to sleep with him" cannot credibly be canceled by "…but only platonically, because he gets cold at night") (see, Horn, 2004). Inter-rater reliability was very high for all three questions, $\alpha = .972$, $\alpha = .905$, and $\alpha = .904$, respectively; the experimenter's ratings were used in all analyses because means cannot be computed to reconcile different ratings for such categorical data.

[10] We did not have enough statistical power to compare rates of discomfort reported by females who made the direct sexual proposition ($n = 19$) and females who received it ($n = 18$), but we note that 47% of female proposers reported it as uncomfortable, while only 33% of female receivers did so.

In response to the question about what was asked at the end of the date, the vast majority of participants reported a direct sexual proposition in the direct condition (87%), an indirect sexual proposition in the indirect condition (87%), and no sexual proposition in the neutral condition (92%) (see Figure 2.1). These percentages were not significantly different, $\chi^2(2, N = 220) = 1.28$, $p = .528$, $\phi = .08$, showing that participants understood the literal denotations of each of the three propositions about equally well.

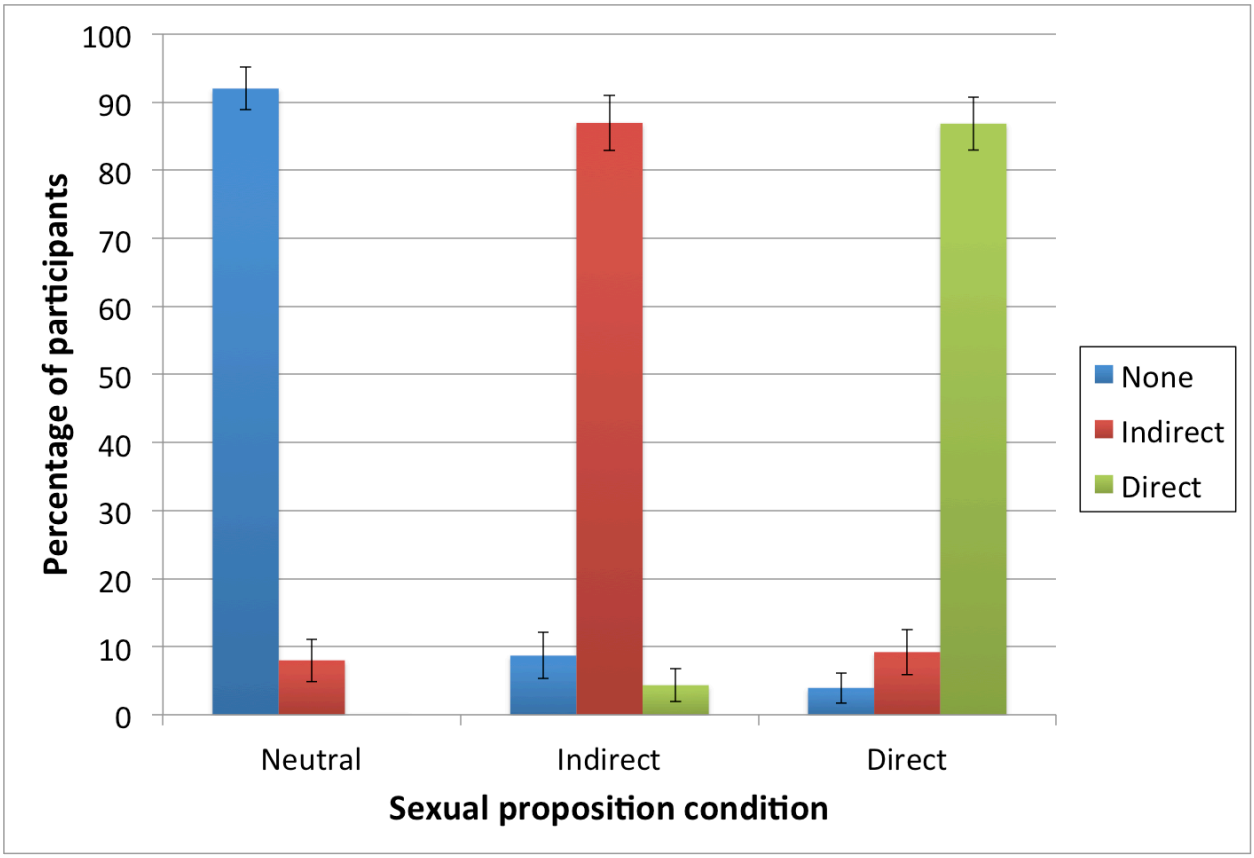

Figure 2.1 Percentage of participants who reported no sexual proposition ("None"), an indirect sexual proposition, or a direct sexual proposition in each date condition, in response to a question about what was asked at the end of the date scenario. Error bars represent standard error.

In response to the question about what the proposer wanted, 54% of the participants in the indirect condition made a direct reference to sex, 26% made an indirect reference to sex, and

only 20% made no reference to sex (see Figure 2.2). This pattern of responses was significantly more similar to the pattern in the direct condition (79%, 8%, and 13%, respectively), $\chi^2(2, N = 145) = 11.81$, $p = .003$, $\phi = .29$, than to the pattern in the neutral condition (11%, 12%, and 77%, respectively), $\chi^2(2, N = 144) = 48.41$, $p < .001$, $\phi = .58$, as shown by a Fisher Z-transformation comparing these two effect sizes, $Z = 3.06$, $p = .002$. Moreover, the percentage of participants who made some kind of reference to sex, either directly or indirectly, was not significantly different between the indirect (80%) and direct (87%) conditions, $\chi^2(1, N = 145) = 1.33$, $p = .248$, $\phi = .10$, suggesting that participants generally understood the implied meaning of the indirect sexual proposition as well as they understood the meaning of the direct proposition.

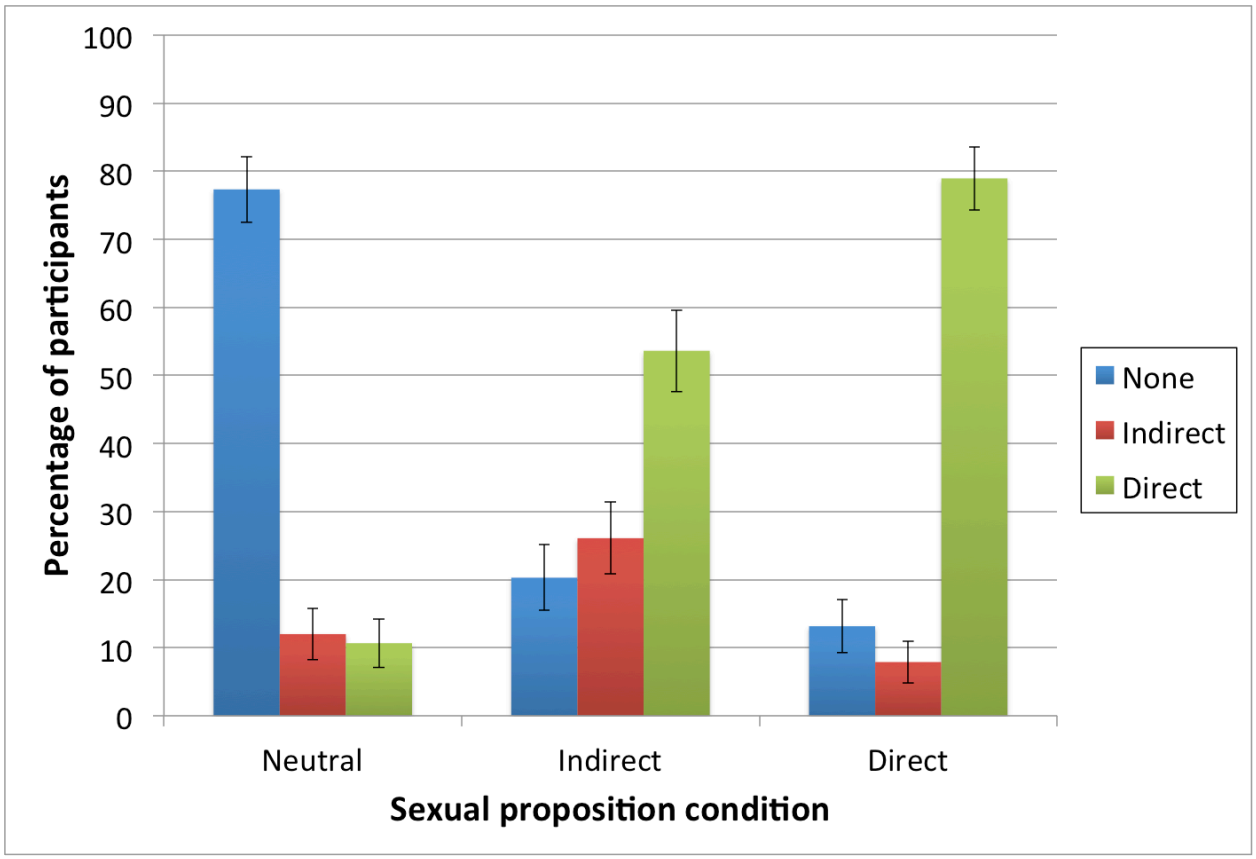

Figure 2.2 Percentage of participants who made no reference to sex ("None"), made an indirect reference to sex, or made a direct reference to sex in each date condition, in response to a question about what the proposer wanted. Error bars represent standard error.

However, when asked to write down what they would tell their best friend the next day about the date, only 8% of participants in the indirect condition made any direct, unambiguous reference to sex (see Figure 2.3). This was comparable to the 6% who inexplicably did so in the neutral condition, $\chi^2(1, N = 135) = 0.31$, $p = .580$, $\phi = .05$, and significantly less than the 53% who did so in the direct condition, $\chi^2(1, N = 137) = 31.28$, $p < .001$, $\phi = .48$. Perhaps more importantly, this 8% was substantially less than the 54% of participants in the indirect condition who made a direct reference to sex in response to the previous question about intent, Wilcoxon Signed-ranks, $Z = 5.12$, $p < .001$, $r = .65$. It's worth noting that this was also true of participants in the direct condition, (53% vs. 79%, respectively), $Z = 2.92$, $p = .003$, $r = .34$; however, a comparison of these two effect sizes using a Fisher Z-transformation revealed that this difference was significantly less than in the indirect condition, $Z = 2.40$, $p = .018$. Lastly, more participants made an indirect reference to sex in the indirect condition (62%) than in either the neutral condition (13%), $\chi^2(1, N = 135) = 35.79$, $p < .001$, $\phi = .51$, or direct condition (20%), $\chi^2(1, N = 137) = 16.88$, $p < .001$, $\phi = .35$, but such references were by definition vague or ambiguous (e.g., "then he asked me if I wanted to come over to his place").

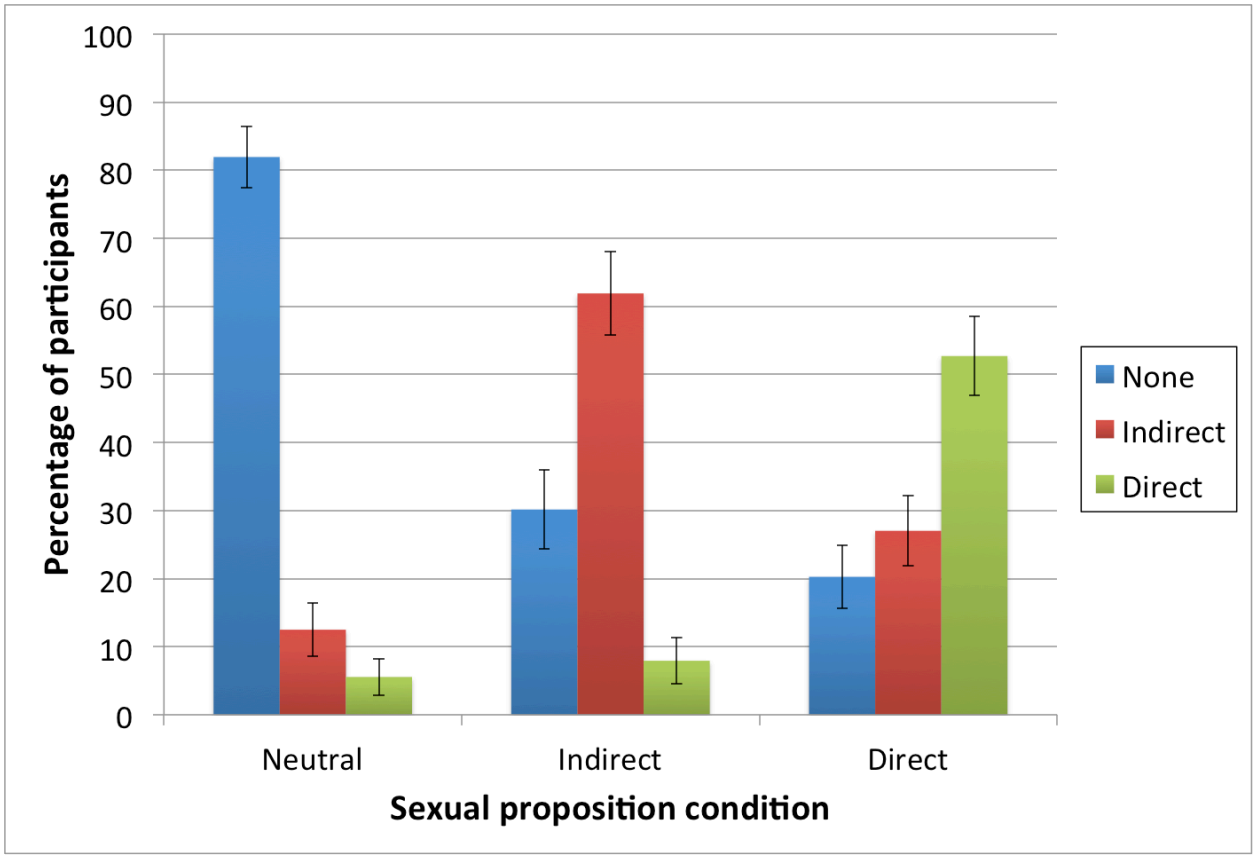

Figure 2.3 Percentage of participants who made no reference to sex ("None"), made an indirect reference to sex, or made a direct reference to sex in each date condition, when asked to write down what they would tell their best friend about the date. Error bars represent standard error.

Cardiovascular Response Measures

Before analyzing any of the cardiovascular reactivity measures, we checked whether the order of the two scenarios interacted with any of the other independent variables in either scenario, which would have made any analyses problematic. This was done for each of the three physiological measures by first running a 3 × 2 × 2 × 2 ANOVA that included all three independent variables of interest—speech condition, role, and sex—as well as order of scenarios. No significant order effects were found, so order was ignored in all subsequent analyses.

Social status. Social status was assessed in two ways, both through a direct self-report measure, and by assaying testosterone levels from saliva samples given at the beginning of the

experiment. Both variables were first centered by subtracting the mean of all scores from each individual participant's score, and then contrasts were run to assess whether either status variable interacted with VC responses in either of the two scenarios (VC is sensitive to general intensity/arousal, so only VC responses were analyzed to minimize the number of comparisons run). Subjective social status was centered based on all participant responses; however, testosterone levels were centered for males and females separately, and contrasts were run for each sex separately, because mean testosterone levels differ substantially between males and females (Mazur & Booth, 1998). Neither subjective social status (all *p*s > .127), nor testosterone levels (all *p*s > .143) showed any significant interactions with VC reactivity in either scenario, and were thus ignored in further analyses.

Date scenario. A 3 × 2 × 2 ANOVA (Speech Condition × Sex × Role) revealed a main effect of speech condition on VC reactivity, $F(2, 174) = 3.19$, $p = .044$, which was qualified by an interaction between sex and speech condition, $F(2, 174) = 3.13$, $p = .046$ (see Figure 2.4). There were no other significant effects.

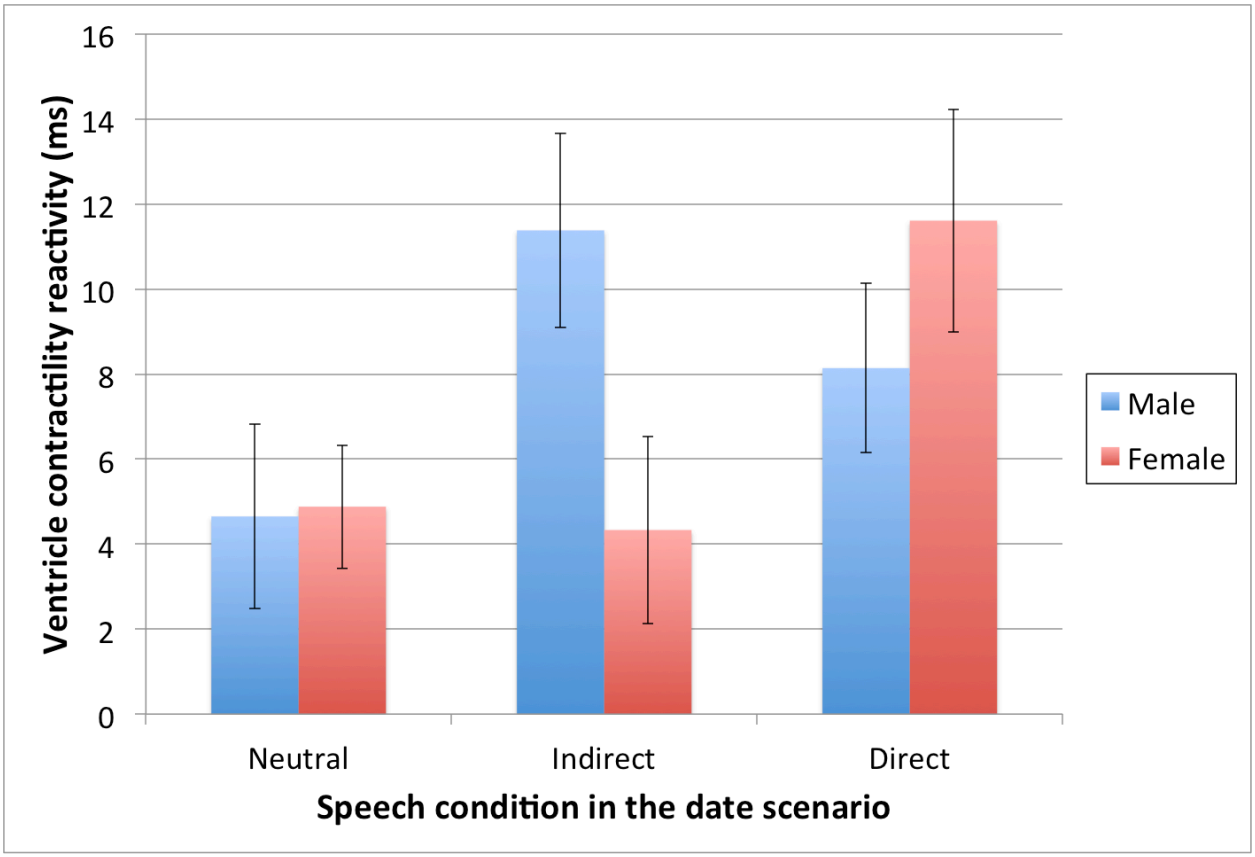

Figure 2.4 Increases in ventricle contractility in the date scenario, organized by sex and speech condition. Error bars represent standard error.

This interaction was decomposed with two simple effects tests comparing the effect of the different speech conditions for males and females. Males showed a marginally significant effect of speech condition on VC, $F(2, 180) = 2.39$, $p = .095$. Follow-up pair-wise comparisons revealed that VC increases were significantly greater for males in the indirect condition than in the neutral condition ($p = .035$), but not in the direct condition ($p = .232$), and there was no significant difference between the indirect and direct conditions ($p = .302$). Females showed a significant effect of speech condition on VC, $F(2, 180) = 3.53$, $p = .031$. Follow-up pair-wise comparisons revealed that VC increases were significantly greater in the direct condition than in both the neutral ($p = .024$) and indirect conditions ($p = .020$); there was no significant difference between the neutral and indirect conditions ($p = .858$). In sum, males had the greatest increase in

ventricle contractility (VC) in the indirect condition, and lowest in the neutral condition, while females showed elevated VC only in the direct condition (see Figure 2.4).

A 3 × 2 × 2 ANOVA (Speech Condition × Sex × Role) revealed no main effect of speech condition on SBP, $F(2, 155) = 1.23$, $p = .295$, but did reveal an interaction between speech condition and sex, $F(2, 155) = 3.83$, $p = .024$. This two-way interaction was further qualified by a three-way interaction between speech condition, sex, and role, $F(2, 155) = 3.10$, $p = .048$ (see Figure 2.5). There were no other significant effects. To better understand the three-way interaction, a 3 × 2 ANOVA (Speech Condition × Sex) was run for each of the two roles. Neither of these tests revealed any main effects; however, proposers showed a significant Speech Condition × Sex interaction, $F(2, 77) = 4.81$, $p = .011$, while receivers did not, $F(2, 78) = .08$, $p = .927$. The two-way interaction in the proposer condition was then further decomposed using simple effects tests that compared the effects of each sexual proposition on male and female proposers separately. These tests revealed a main effect of speech condition for female proposers, $F(2, 77) = 4.77$, $p = .011$, but not for male proposers, $F(2, 77) = .99$, $p = .376$. Follow-up pairwise comparisons revealed that female proposers had a greater SBP increase in the direct condition than the neutral condition ($p = .016$), but not than the indirect condition ($p = .252$), and there was no significant difference between the indirect and neutral conditions ($p = .176$). To summarize, as receivers of the proposition, neither sex showed any significant difference in SBP responses across the three speech conditions (see Figure 2.5a); however, female proposers had a significantly greater increase in SBP in the direct condition, while male proposers showed no significant difference across the three speech conditions (see Figure 2.5b).

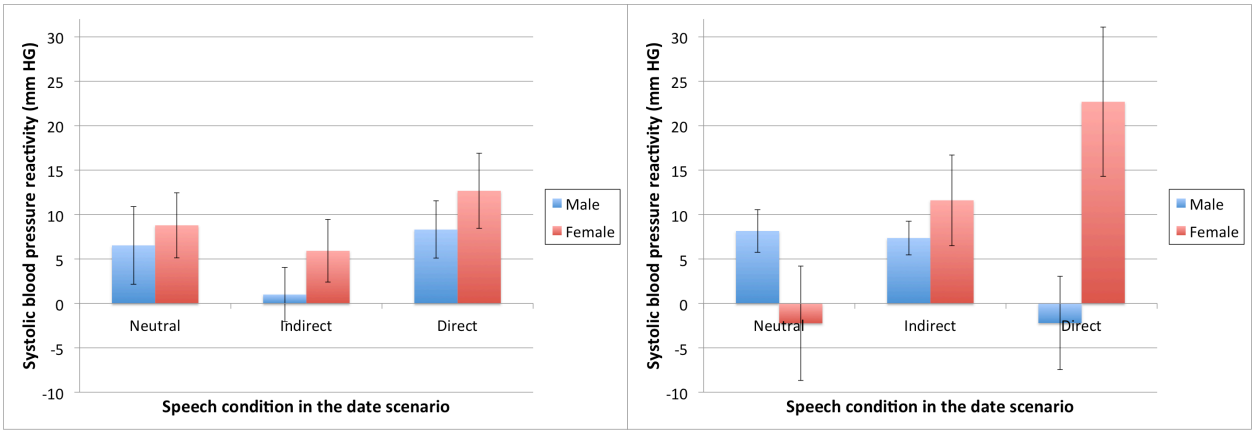

Figure 2.5 Increases in systolic blood pressure in the date scenario, organized by sex and speech condition, and in separate figures for receivers (2.5a, on the left) and proposers (2.5b, on the right). Error bars represent standard error.

No significant effects were found for CO in the date scenario.

Restaurant scenario. A 3 × 2 × 2 ANOVA (Speech Condition × Sex × Role) revealed a main effect of speech condition on VC, $F(2, 168) = 3.75$, $p = .026$ (see Figure 2.6). There were no other significant effects. Post hoc tests showed significantly greater VC in the direct condition than in either the neutral ($p = .026$), or indirect conditions ($p = .013$), and no significant difference between the indirect and neutral conditions ($p = .823$). Participants exhibited greater sympathetic activation to the direct bribe proposition, than to either of the other two propositions.

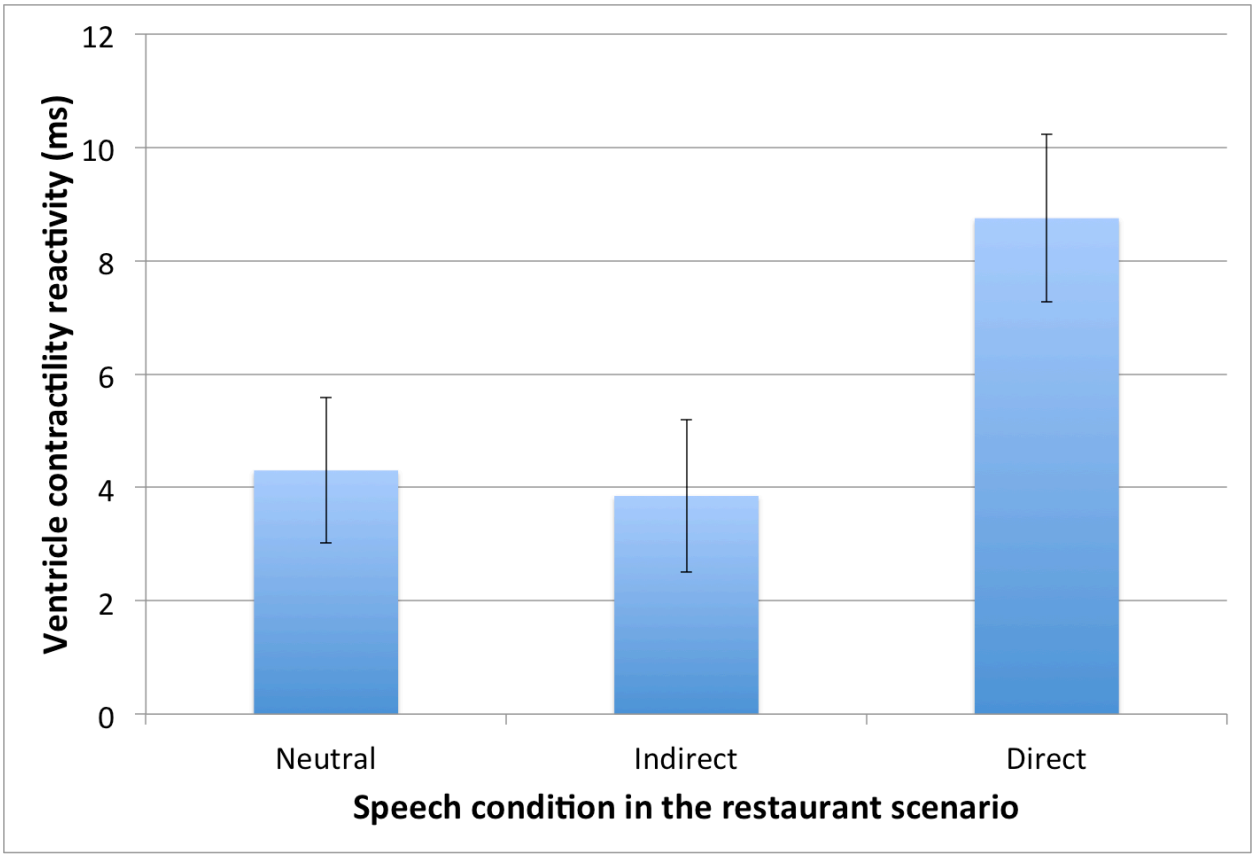

Figure 2.6 Increases in ventricle contractility by speech condition in the restaurant scenario. Error bars represent standard error.

A 3 × 2 × 2 ANOVA (Speech Condition × Sex × Role) revealed no main effects on CO, but did reveal a Speech Condition × Role interaction, $F(2, 166) = 3.88$, $p = .023$ (see Figure 2.7). There were no other significant effects.

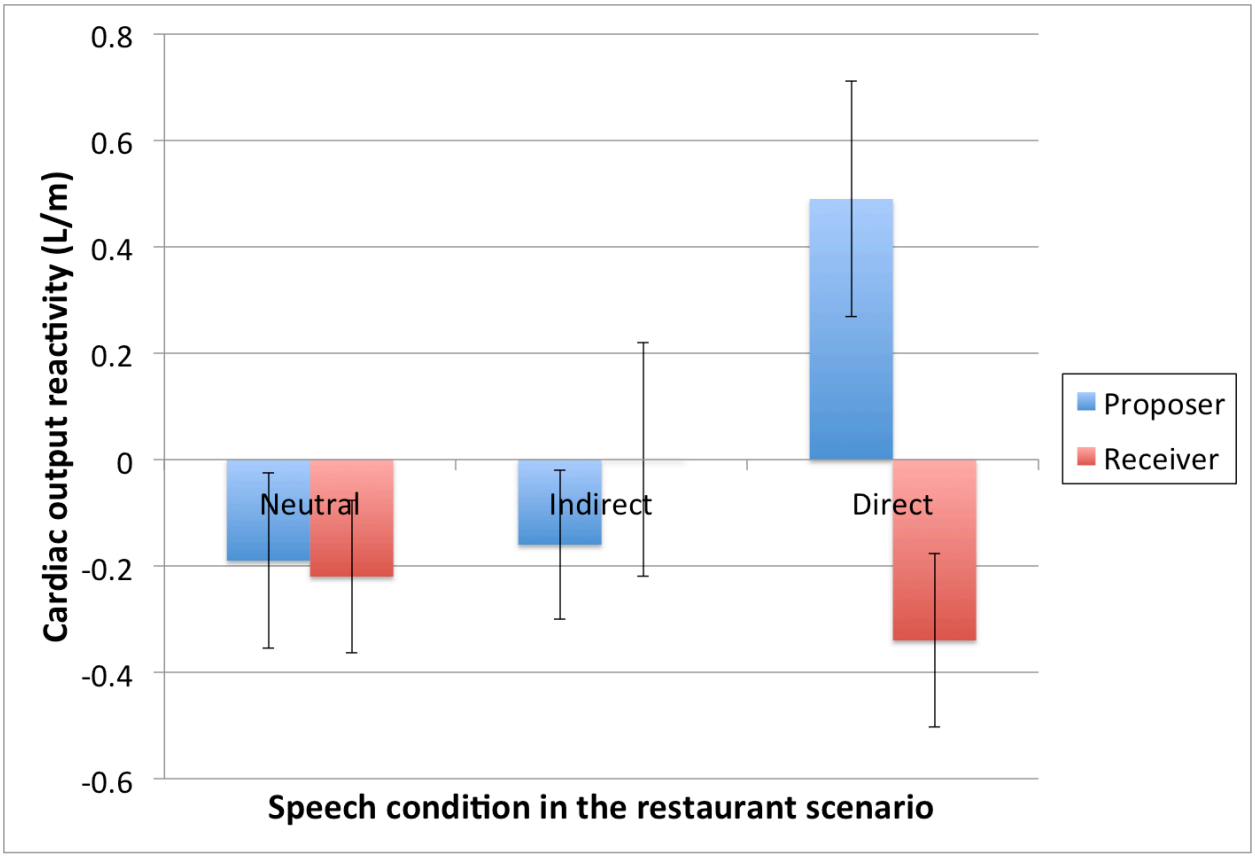

Figure 2.7 Increases in cardiac output in the restaurant scenario, organized by speech condition and role. Error bars represent standard error.

This interaction was decomposed with simple effects tests comparing speech conditions for proposers, whose CO differed across conditions, $F(2, 85) = 4.49$, $p = .014$, and receivers, whose CO did not, $F(2, 87) = 0.92$, $p = .402$. Follow-up pair-wise comparisons revealed that proposers' CO increased more in the direct condition than in the indirect ($p = .014$) and neutral conditions ($p = .011$), with no difference between the indirect and neutral conditions ($p = .933$). In sum, participants that offered the bribe (patrons) showed increases in CO only in the direct condition, while participants who received the bribe (maitre d's) did not show increased CO in any condition (see Figure 2.7).

No significant effects were found for SBP in the restaurant scenario.

Discussion

This study provides the first evidence that indirect and direct propositions lead to different affective states, as hypothesized by the strategic speaker theory (SST), and the results suggest that affective reactions track costs and benefits in a manner consistent with the SST. However interpreting the story these data tell is complex, because even though the overall pattern of results was broadly consistent with the theory, the specific affective states resulting from the two types of propositions did not conform to our specific predictions. Consistent with the theory, participants' cardiovascular reactivity profiles to direct and indirect propositions were different in both scenarios. Furthermore, male and female response profiles were different in the date scenario, while proposers and receivers exhibited different response profiles in the restaurant scenario, as predicted by how each of these variables uniquely affects the costs and benefits of each scenario. However, direct and indirect propositions did not consistently lead to threat and challenge states, respectively, as we had predicted. Despite this, we nonetheless think that the overall pattern of results supports the SST.

In the date scenario, females' cardiovascular profiles were consistent with a threat response (increased VC and SBP) when they made the direct sexual proposition; no difference was found between their responses in the indirect and neutral conditions. Males' cardiovascular responses to both sexual propositions, both as proposers and receivers, were consistent with a challenge response (increased VC, no increase in SBP), with the strongest response (greatest increase in VC) in the indirect condition. In the restaurant scenario, participants who made the direct bribe proposition (patrons) exhibited cardiovascular profiles consistent with a challenge response (increased VC and CO). Participants who received the direct bribe offer (maitre d's) showed cardiovascular profiles consistent with weaker challenge responses, as indicated by the

significant difference in CO responses between patrons and maitre d's. No differences were found between the indirect and neutral conditions for either patrons or maitre d's.

Despite these differences in cardiovascular responses observed across speech conditions in both scenarios, no differences were found in self-reported positive or negative affect, affirming the supposition that the affective concomitants of indirect and direct speech are difficult to assess in the lab.

The general prediction that direct propositions would elicit threat responses and indirect propositions would elicit challenge responses did not bear out. However, the rates of reported discomfort across conditions suggest that this was because participants did not consistently perceive the costs of the direct propositions to outweigh the benefits. More than six times as many participants reported that the date scenario made them uncomfortable than the restaurant scenario, and rates of reported discomfort did not differ across speech conditions in the restaurant scenario. Females in the direct sexual proposition condition showed the highest rates of reported discomfort of any group, while males' rates of reported discomfort were not significantly different across the date conditions. These results mirror the cardiovascular results, as the only group that exhibited threat responses—females who made the direct sexual proposition—was also more likely than any other group to report that the interaction made them uncomfortable. This suggests that participants did not generally perceive the costs of the direct bribe proposition to outweigh its benefits, nor did males generally perceive the costs of the direct sexual proposition to outweigh its potential benefits. Thus, even though some of the cardiovascular results were inconsistent with the specific affective states we predicted, they nonetheless seem to be consistent with the SST—which predicts that people should make direct propositions when they perceive the benefits to outweigh the costs, because indirect propositions

may be misunderstood (see Lee & Pinker, 2010)—and may simply be the result of weaknesses in our experimental design.

A number of limitations make these interpretations far from conclusive. First, the cardiovascular results did not provide complete challenge and threat profiles, which makes it difficult to precisely and confidently infer affective states. The cardiovascular profiles were incomplete both because participant movement precluded reliable TPR estimates, and because neither scenario produced significant differences for all three cardiovascular measures associated with challenge and threat responses (VC, CO, and TPR, approximated here with SBP). Second, the metric used to interpret how participants perceived costs and benefits—reported discomfort—was indirect and imprecise, and the interpretations of the cardiovascular results based on this metric were necessarily post hoc. Third, some of the cardiovascular results were only marginally or barely statistically significant, and the statistical analyses required a number of different comparisons.

Despite these limitations however, we find the above interpretations convincing because the overall pattern of results largely align with clear predictions from the SST. First, participants' cardiovascular responses were different in the direct and indirect conditions in both scenarios, even though the response profiles were not exactly as predicted in terms of challenge and threat. Second, sex was the main moderating variable in the date scenario, and role was the only moderating variable in the restaurant scenario, a pattern clearly predicted a priori by how each of these two variables uniquely affects the costs and benefits specific to each scenario. Third, even though the alignment of reported discomfort with cardiovascular responses only allowed for a post hoc interpretation of unexpected results, it seems unlikely that such an alignment would have occurred solely by chance. Fourth, the contrived, scripted interactions used here provide a

conservative and weak test of how people might react to such situations in natural environments with real social consequences (see Feiler, 2000, for an account of how trying to bribe a maitre d' can be much more uncomfortable than the rates reported here would suggest). The divergence between participants' self-reported affect and affective responses inferred from cardiovascular responses, suggests that such an artificial, contrived set up was nonetheless justified, because it allowed us to bring the phenomenon into the lab. As such, these interpretations seem both warranted and valid, but clearly more research is needed to further substantiate the affective predictions of the SST.

The results also supported the prediction that indirect speech can be used to prevent rumors. Participants clearly reported what was said in the different date scenario conditions, and demonstrated that they generally understood the intended meaning of the indirect sexual proposition, yet they only reported that they would tell their best friend about the sexual proposition in an unequivocal way when it was expressed directly. Even though participants who role-played the indirect sexual proposition were most likely to make an indirect reference to sex when recounting the date to their friend, these reports were often ambiguous and unclear (e.g., "then he asked me if I wanted to come over to his place"). Furthermore, the difference between the majority of participants who directly reported the intended meaning of the indirect sexual proposition to us, and the small percentage who said they would directly report it to their friend, suggests that such gossip would be substantially less clear and unambiguous than participants' own understanding of what was communicated. This conclusion depends on two assumptions: (1) People actually report indirect propositions indirectly to third parties in real world situations, as observed here; and, (2) Recipients of such indirect reports are less than certain about the speaker's intended meaning. Further research is needed to substantiate these assumptions.

General Discussion and Conclusion

By adding to the sparse extant evidence for the SST, this study helps establish an empirical foundation for a more comprehensive understanding of indirect speech, which extends beyond previous explanations predicated on purely cooperative interactions. We suggest however, that the implications of this research are much broader than just explaining why people use one form of speech or another. Two core insights of the SST—that relationship negotiation is a kind of coordination game, and that this endows common knowledge with a uniquely important role in navigating relationships—apply to many different kinds of social interactions, not just those between conversational partners. By providing evidence for some of the SST's predictions about how people respond to different kinds of propositions, this study also provides indirect evidence for these much broader insights, which may help explain many other aspects of human social life.

Coordination can be facilitated by many different kinds of communication, knowledge can be gleaned from many diverse types of information, and human social life provides numerous contexts in which the outcome of some action or event that everyone privately knows about can be dramatically different than if it becomes common knowledge. The well-documented *omission bias* in moral cognition provides one illustrative example. Moral transgressions are punished more harshly and judged as more wrong, when they result from an action (a *commission*) than from a failure to act (an *omission*), even when the intent and outcome are the same (e.g., poisoning someone is judged less harshly than withholding a readily available antidote, Cushman, Young, & Hauser, 2006). One recent explanation of this phenomenon suggests that this is because omissions leave no physical evidence, which precludes common knowledge of the transgression (DeScioli & Kurzban, 2013). Consistent with this explanation,

when omissions leave behind some kind of evidence that can be made common knowledge, they are judged and punished just as harshly as commissions (DeScioli, Bruening, & Kurzban, 2011). Similarly, people judge unethical propositions as less morally wrong when they are delivered indirectly than directly, and unethical behaviors described with transparent euphemisms are judged as less morally wrong and less deserving of punishment than direct descriptions of the same behaviors (Chakroff, Thomas, Haque, & Young, 2014). Integrating these two parallel lines of research suggests that the coordination dynamics invoked by the SST may also explain the difference between moral judgments and punishment of omissions and commissions: Omissions are akin to indirect propositions, while commissions are akin to direct propositions.

The omission bias offers a convenient example because such parallel lines of research already exist, but we suggest that a number of other social phenomena may also be explained by the dynamics outlined in the SST. For example, Ari Adut (2008) has offered a similar explanation for the sociological phenomenon of scandals, and has marshaled an impressive body of evidence to show that scandals often occur when something that was previously known and discussed by many people in private, suddenly becomes public common knowledge. Rather than working through the details of many more similar examples that seem to share these dynamics, here we simply list some of the many possibilities (see Lee & Pinker, 2010, and Thomas, et al., 2014 for further elaboration): taboo, conventional wisdom, diffusion of responsibility, etiquette, celebrity, tact, fame, hypocrisy, mock outrage, arbitrary social conventions, pluralistic ignorance, political correctness, double speak, currency valuation, negotiation and bargaining, the creation and popping of market bubbles, and international relations and diplomacy. We hope that future research can explore, clarify, and substantiate whether these seemingly diverse phenomena might be explained by similar strategic dynamics.

PAPER 3

TWO KINDS OF PUBLIC: COMMON KNOWLEDGE AMPLIFIES NEGATIVE SELF-CONSCIOUS EMOTIONS

Kyle A. Thomas[1], Peter DeScioli[2], Steven Pinker[1]

[1]Department of Psychology, Harvard University

[2]Department of Political Science, Stony Brook University

Abstract

People's social transgressions elicit the negative self-conscious emotions of embarrassment, shame, and guilt. Previous research shows that the presence of an audience intensifies these emotions. However, they seem to be even more intense when the actor acknowledges their transgression with the audience. This difference between a person's transgression being acknowledged and being observed parallels a distinction in game theory between *common knowledge* (an infinite recursion of shared knowledge states, such that A knows that B knows that A knows that B knows, ad infinitum), and *shared knowledge* (any lesser, finite level of shared knowledge states). We test whether self-conscious emotions are sensitive to this distinction. In Experiment 1, participants read vignettes and reported how they would feel in emotion-evoking scenarios with varying levels of knowledge. In Experiment 2, participants reported their emotions after performing a karaoke song for a panel of judges over a live two-way video feed, with either shared or common knowledge. Results from both experiments showed that self-conscious emotions were more intense when the focal event was common knowledge than when it was shared knowledge. The results show that not only the presence of an audience, but also the precise level of knowledge one shares with the audience affects the intensity of self-conscious emotions. We discuss how the regulatory functions of self-conscious emotions—managing relationships and social status—can be illuminated by game theory models in which players' best strategies depend on whether they have shared versus common knowledge.

Introduction

Imagine tripping over your feet in front of a crowd. Afterwards, you might fix your gaze on your cell phone or a magazine to avoid eye contact with the audience. Similarly, after disappointing your family, friends, or coworkers, it can be hard to look them in the eye. Why do people avoid acknowledging faux pas or transgressions that they know an audience already knows about?

Transgressions and faux pas evoke the negative self-conscious emotions of shame, embarrassment, and guilt, which help people manage relationships and social status (Beer, Heerey, Keltner Scabini, & Knight, 2003; Keltner & Buswell, 1997; Miller, 1995; Steckler & Tracy, 2014; Tangney & Tracy, 2012; Tracy & Robins, 2004). People's transgressions can harm their reputation or relationships if other people know about them, and research confirms that the presence of an audience intensifies feelings of embarrassment, shame, and guilt (Bosch et al., 2009; Gruenewald, Kemeny, Aziz, & Fahey, 2004; Modigliani, 1971; Smith, Webster, Parrot, & Eyre, 2002; Tangney, Miller, Flicker, & Barlow, 1996; Wolf, Cohen, Panter, & Insko, 2010). However, these studies did not test whether these feelings are more intense when an actor acknowledges their offense.

This difference between evading and acknowledging unflattering events parallels a distinction in game theory between *common knowledge* and *shared knowledge*. Common knowledge is defined as an infinite recursion of shared knowledge states, such that A knows X, A knows that B knows X, A knows that B knows that A knows X, A knows that B knows that A knows that B knows X, ad infinitum. In contrast, shared knowledge refers to any lesser, finite level of such knowledge states (e.g., A knows that B knows that A knows X, but nothing more). Game theory models show that in many situations, a person's best move differs depending on

whether they have common knowledge or shared knowledge about the payoffs (Chwe, 2001; Rubinstein, 1989; Schelling, 1960). Further, psychological research shows that people do in fact represent common knowledge and shared knowledge as qualitatively distinct and they make different strategic decisions based on this distinction, consistent with game theoretic models (Thomas et al., 2014).

If the self-conscious emotions have strategic functions, then the difference between common and shared knowledge might explain why people avoid eye contact after a faux pas or transgression. Specifically, people's evasive behaviors might function to prevent common knowledge. Moreover, people's intensified self-conscious emotions might alert them when common knowledge of a transgression occurs so they can manage its strategic consequences, or motivate them to avoid such occurrences in the future.

The Present Research

In two experiments we investigate whether the negative self-conscious emotions are sensitive to the difference between common knowledge and shared knowledge. In Experiment 1, participants read vignettes describing situations that typically evoke either negative self-conscious emotions or the negative basic emotions of sadness or anger, with varying levels of knowledge. Participants then rated how intensely they would expect to feel these emotions and exhibit associated physical responses. The strategic hypothesis predicts that participants will report more intense self-conscious emotions when an event is common knowledge than when it is shared knowledge.

Experiment 2 tests whether knowledge levels and eye contact affect actual experiences of embarrassment. Participants performed a karaoke song for a panel of judges in a separate room connected by a live two-way video feed, and then rated how strongly they felt a set of emotions.

In the common knowledge condition, participants were told that the judges knew about the two-way live video feed. In the shared knowledge condition, participants were told that the judges falsely believed they were watching a one-way video feed that was unknown to participants. In addition, the video of the judges that participants saw either showed the judges looking straight into the camera or looking slightly to the side. The strategic hypothesis predicts that common knowledge will lead to greater feelings of embarrassment. Further, if eye contact is a reliable cue for common knowledge, then the presence of eye contact should similarly elicit more intense embarrassment.

Experiment 1

In Experiment 1 participants reported how they would react to different levels of knowledge in five scenarios. Three of the scenarios were designed to evoke negative self-conscious emotions. Participants imagined that they got caught making fun of a friend, audibly passing gas during a lecture, or dishonestly reporting purchases to a club for reimbursement. Two basic emotion scenarios were included as controls to test whether any effect of different knowledge levels is specific to self-conscious emotions. In one scenario, designed to evoke sadness, participants imagined discovering their romantic partner was about to end the relationship. In another scenario, designed to evoke anger, participants imagined discovering an acquaintance trying to sabotage a valued friendship.

To test the effect of knowledge levels on reported emotions, each participant read multiple vignettes that described different levels of knowledge of the focal event in one of the five scenarios. That is, the scenario varied between-subjects whereas knowledge levels varied within-subjects. The different knowledge levels included *private knowledge*, in which participants shared no knowledge of the focal event with another person, *shared knowledge*, in

which participants and another person had some finite level of shared knowledge about the focal event, and *common knowledge*, in which the focal event was commonly known between participants and another person.

Because negative self-conscious emotions are typically caused by one's own transgressions, the focal event in the self-conscious scenarios involved something the participant did that could be discovered by an observer. However, because negative basic emotions tend to be evoked by other people's actions, the focal event in the basic emotion scenarios involved the participant as an observer discovering something about someone else's behavior. That is, participants assumed the role of the *actor* in the self-conscious scenarios and the role of an *observer* in the basic emotion scenarios.

Method

Participants. One hundred participants from the United States were recruited from Amazon's Mechanical Turk service for each of the five scenarios ($N = 500$), and paid $2-$3 for their participation. Participants who failed comprehension checks or an attention check (see Procedure) were excluded, yielding a final sample of 361 participants (55% female, $M_{age} = 34.1$, $SD_{age} = 11.1$).

Procedure. Participants read instructions explaining the task and were given definitions of the six emotions, derived from standard dictionaries, that they would be asked to rate in the study:

- *Anger*—A strong feeling of displeasure, grievance, annoyance, and hostility
- *Embarrassment*—A feeling of acute and painful self-consciousness, uncomfortableness, awkwardness, and humiliation

- *Fear*—An unpleasant feeling of anxiety or apprehension caused by the presence or anticipation of danger
- *Guilt*—An awareness of having done wrong or failed in an obligation, accompanied by feelings of remorse and regret
- *Sadness*—A feeling of unhappiness, grief, and sorrow
- *Shame*—A painful negative emotion that combines feelings of dishonor, unworthiness, distress, and humiliation, and is caused by the awareness of having done something wrong or foolish

To ensure they clearly understood how each emotion is characterized, participants then answered six comprehension questions in which they were given a definition and had to select the corresponding emotion. Participants who answered any of these questions incorrectly were excluded from analyses.

Participants then read five versions of one scenario, each with a different level of knowledge (presented in a random order). After each vignette, participants used slider bars to report how strongly they would experience each of the six emotions on a scale of 0-100, and how likely and intensely they would exhibit 16 different physical reactions associated with these different emotions on a scale of 0-100. These physical responses were drawn from previous emotion research and included: avoid gaze, blanch, blush, clench teeth, cover or touch face, cry, frown, glare, hang head down, nervous laughter, open eyes wide, raise eyebrows, sheepish grin, slump shoulders, stammer, and tremble (see Ekman, 2003; Keltner & Buswell, 1997; Tracy & Robins, 2004).

Participants then reported basic demographic information and responded to a simple attention check. This attention check was disguised as a final question that asked participants to

report which emotions from a list of 20 they were experiencing but had an instruction buried in the prompt that asked them to select "only the enthusiastic option" to show that they were paying attention. Participants who did not select this option were excluded from analyses. Participants were then debriefed and paid.

The five scenarios. Each participant read vignettes describing one of the following five scenarios which were designed to elicit self-conscious emotions (teasing, farting, and dishonesty), sadness (break up), or anger (sabotage):

- Self-conscious emotion scenarios
    - *Teasing*—Participants took the perspective of someone talking to a friend, and mocking a mutual friend's speech impediment, when the person being mocked unexpectedly shows up.
    - *Farting*—Participants took the perspective of someone who audibly farts in a crowded lecture hall.
    - *Dishonesty*—Participants took the perspective of a person planning to submit a dishonest reimbursement request to a club, which is discovered by their housemate.
- Basic emotion scenarios
    - *Break up*—Participants took the perspective of someone who discovers that their romantic partner is going to break up with them by overhearing a voicemail left by their romantic partner's friend.
    - *Sabotage*—Participants took the perspective of someone who overhears an acquaintance who is trying to sabotage their relationship with a good friend.

The five knowledge levels. Each participant read five different versions of a scenario with different levels of knowledge about the focal event. Because participants assumed the role of the actor in the self-conscious scenarios, but assumed the role of observer in the basic emotion scenarios, participants had different kinds of knowledge in some of the conditions in the two types of scenarios. As such, the knowledge levels are broken out separately here for the self-conscious emotion scenarios and basic emotion scenarios, and labeled according to the level of knowledge participants had.

- Self-conscious emotion scenarios:
    1. *Private knowledge 1*—Only the participant knows about their transgression
    2. *Private knowledge 2*—An observer is aware of the participant's transgression, but the participant doesn't realize this
    3. *Shared knowledge 1*—The participant knows that an observer knows about their transgression, but the observer is unaware they know this
    4. *Shared knowledge 2*—The participant knows that an observer knows about their transgression, the observer knows the participant knows this, but the observer is unaware that the participant knows they know this
    5. *Common knowledge*—The participant's transgression is commonly known (i.e., the participant knows the observer knows about it, knows the observer knows they know, ad infinitum)
- Basic emotion scenarios:
    1. *No knowledge*—The participant is unaware of the focal event
    2. *Private knowledge 1*—The participant is aware of the focal event, but the other person doesn't realize this

3. *Private knowledge 2*—The participant is aware of the focal event, and the other person knows this, but the participant doesn't know the other person is aware they know
4. *Shared knowledge*—The participant is aware of the focal event, and the participant knows that the other person is aware that they know about it, but the other person is unaware they know this
5. *Common knowledge*—The focal event is commonly known (i.e., the other person knows the participant knows about it, knows the participant knows they know, ad infinitum)

The full text of all vignettes is supplied in Appendix B of the Supplementary Materials.

Data analysis. Although shame, embarrassment, and guilt are distinct emotions, they are similar enough that people often conflate them (Keltner & Buswell, 1996; Tangney et al., 1996; Tangney & Tracy, 2012). Thus, as a preliminary step, we analyzed correlations between these emotions within each knowledge condition of the three self-conscious emotion scenarios. As expected, ratings of embarrassment, shame, and guilt were highly correlated in every knowledge condition of all three self-conscious scenarios, so scale scores were computed by averaging ratings of these emotions within each of the five knowledge conditions of all three scenarios (reliabilities of these 15 *SCE* scale scores ranged from $\alpha = .791$ to $\alpha = .948$). Similarly, self-reported physical reactions associated with the self-conscious emotions—blush, cover or touch face, hang head down, nervous laughter, sheepish grin, slump shoulders, and stammer[11]—were also highly correlated, so analogous scale scores were computed (reliabilities of these 15 scale

[11] Gaze avoidance was not included in these composites, even though it is commonly associated with the self-conscious emotions, because avoiding or making eye contact can alter the prevailing level of knowledge.

scores ranged from α = .682 to α = .850). These scale scores are used for all analyses of the self-conscious emotions and their physical concomitants in the self-conscious emotion scenarios.

For each scenario, we first perform manipulation checks to test whether each scenario evoked the intended emotion by comparing the mean reported level of each emotion, averaged across all five knowledge levels within each scenario[12]. Subsequent analyses then focus on the dominant emotion (the one with the highest mean) and its physical concomitants.

Results

Preliminary analyses. Paired sample t-tests confirmed that ratings of the self-conscious emotions were higher than the next highest rated emotion in the teasing (M_{SCE} = 42.3 vs. M_{sad} = 18.0, $t(77) = 14.71$, $p < .001$, $d = 1.69$), farting (M_{SCE} = 24.8 vs. M_{fear} = 9.4, $t(69) = 11.57$, $p < .001$, $d = 1.38$), and dishonesty scenarios (M_{SCE} = 31.1 vs. M_{fear} = 22.1, $t(64) = 4.58$, $p < .001$, $d = 0.57$). Analogous tests confirmed that ratings of sadness were highest in the break up scenario (M_{sad} = 54.8 vs. M_{anger} = 40.3, $t(76) = 6.44$, $p < .001$, $d = 0.74$), and that ratings of anger were highest in the sabotage scenario (M_{anger} = 52.2 vs. M_{sad} = 21.3, $t(70) = 10.26$, $p < .001$, $d = 1.22$).

Self-conscious emotion scenarios. Figure 3.1 shows self-conscious emotion ratings (Figure 3.1a) and associated physical reactions (Figure 3.1b) across knowledge levels, in each of the three self-conscious scenarios. The results show that self-conscious emotions and their physical concomitants were lowest in the two private knowledge conditions, intermediate in the two shared knowledge conditions, and highest in the common knowledge condition in all three scenarios, as predicted.

[12] All mean emotion ratings for each knowledge condition in every scenario are provided in Table SM1 in Appendix A of Supplementary Materials.

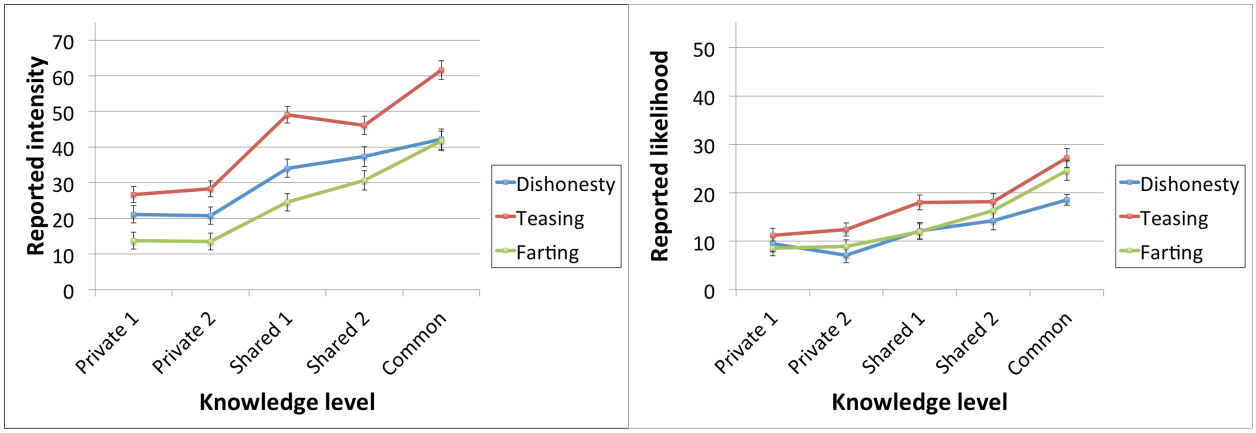

Figure 3.1 Ratings of self-conscious emotions (3.1a, on the left) and likelihood of displaying their physical concomitants (3.1b, on the right) in the three self-conscious emotion scenarios, across the different knowledge levels. Each line in Figure 3.1a represents the combined average ratings of embarrassment, shame, and guilt in the respective scenarios. Each line in Figure 3.1b represents the combined average ratings of seven physical reactions associated with these emotions in the respective scenarios. Error bars represent standard error.

Repeated-measures ANOVAs confirmed that self-conscious emotions significantly increased as knowledge levels increased in the dishonesty, $F(4, 256) = 49.02$, $p < .001$, $\eta_p^2 = .43$, teasing, $F(4, 308) = 53.16$, $p < .001$, $\eta_p^2 = .41$, and farting scenarios, $F(4, 276) = 86.65$, $p < .001$, $\eta_p^2 = .56$. Planned post hoc tests confirmed that in all three scenarios self-conscious emotions were higher in the common knowledge condition than in the shared knowledge conditions (all *p*s < .001), and higher in the shared knowledge conditions than in the private knowledge conditions (all *p*s < .001). Repeated-measures ANOVAs also confirmed that the physical reactions associated with self-conscious emotions significantly increased as knowledge levels increased in the dishonesty, $F(4, 256) = 23.12$, $p < .001$, $\eta_p^2 = .27$, teasing, $F(4, 308) = 36.95$, $p < .001$, $\eta_p^2 = .32$, and farting scenarios, $F(4, 276) = 51.08$, $p < .001$, $\eta_p^2 = .43$. Planned post hoc tests confirmed that in all three scenarios ratings of these reactions were higher in the common knowledge condition than in the shared knowledge conditions (all *p*s < .001), and higher in the

shared knowledge than private knowledge conditions (all *p*s < .005, except $p = .016$ for shared knowledge 1 vs. private knowledge 1 in the dishonesty scenario).

Basic emotion scenarios. Figure 3.2 shows ratings of sadness in the break up scenario and anger in the sabotage scenario across knowledge levels (Figure 3.2a), and the corresponding physical responses of crying and teeth clenching, respectively (Figure 3.2b). The results show that acquiring private knowledge caused a large increase in reported anger and sadness, but that these emotions and their associated physical responses were relatively more stable across the additional levels of knowledge; however, reported anger was higher in the common than shared knowledge condition, and participants reported that they were less likely to cry in the common than shared and private knowledge conditions.

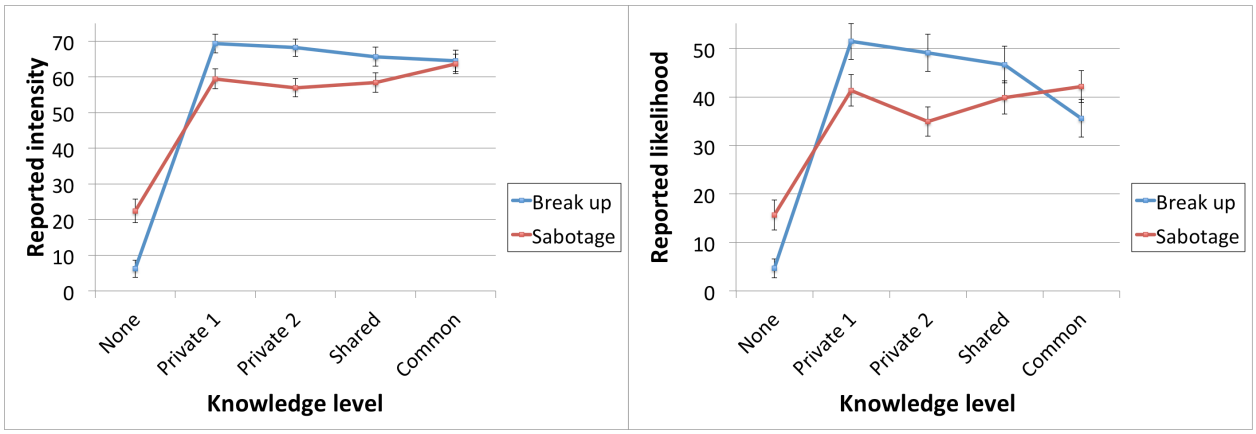

Figure 3.2 Ratings of sadness in the break up scenario and anger in the sabotage scenario (3.2a, on the left), and ratings of likelihood of crying in the break up scenario and teeth clenching in the sabotage scenario (3.2b, on the right), across the different knowledge levels. Error bars represent standard error.

Break up scenario. Repeated-measures ANOVAs revealed significant differences across knowledge levels for both sadness, $F(4, 304) = 203.01$, $p < .001$, $\eta_p^2 = .73$, and the associated physical reaction of crying, $F(4, 304) = 72.33$, $p < .001$, $\eta_p^2 = .49$. However, post hoc tests showed that this was primarily because of the large difference between responses in the no

knowledge condition and the other four conditions, in which responses were all similar. Reported sadness in the no knowledge condition was more than ten-times lower than in the four conditions in which participants had any knowledge of the impending break up (all *p*s < .001), and there was only one slightly significant difference between any of these four conditions ($p = .047$). Participants also reported that they were more than seven-times more likely to cry in the four conditions in which they knew about the impending break up than in the no knowledge condition (all *p*s < .001). The only significant difference between these other four conditions was that participants reported they would be less likely to cry in the common knowledge condition than in the private or shared knowledge conditions (all *p*s < .001).

Sabotage scenario. Repeated-measures ANOVAs revealed significant differences across knowledge levels for both anger, $F(4, 280) = 86.90$, $p < .001$, $\eta_p^2 = .55$, and the associated physical reaction of teeth clenching, $F(4, 280) = 34.42$, $p < .001$, $\eta_p^2 = .33$. Post hoc tests showed that reported anger in the no knowledge condition was more than two-times lower than in the four conditions in which participants had any knowledge (all *p*s < .001), and it was not significantly greater in the shared knowledge condition than in the private knowledge conditions. Reported anger was however significantly greater in the common knowledge condition than in the shared knowledge condition ($p = .004$). Participants also reported that they would be more than twice as likely to clench their teeth in the four conditions in which they knew about the sabotage attempt than in the no knowledge condition (all *p*s < .001), and participants were not significantly more likely to report that they would clench their teeth in the common knowledge condition than in the shared knowledge condition. Oddly, participants reported a significantly lower likelihood of clenching their teeth in the private knowledge 2 condition than in the

common knowledge condition (p = .015), and even than in the private knowledge 1 condition (p = .009).

Experiment 2

The results of Experiment 1 support the strategic hypothesis in hypothetical scenarios. Experiment 2 brings the phenomenon into the lab to test whether the difference between shared and common knowledge affects actual experiences of embarrassment. In Experiment 2, participants performed a karaoke song for a panel of four judges in a separate room through a live two-way video feed and then reported how strongly they felt different emotions. During their performance, participants were shown a karaoke-style video with the lyrics to the song; however, this was embedded in a larger video display showing a live feed of the judges in the other room which forced participants to view the judges for their entire performance (see Figure 3.3). In fact, there were no live judges and participants were shown one of two pre-recorded videos. One video showed the judges looking straight into the camera (Figure 3.3a), whereas the other showed them looking slightly to the side (Figure 3.3b). Additionally, participants were either told that the judges knew that there was a live two-way video feed, and that the judges knew participants were aware they knew this, or were told that the judges falsely believed they were watching a one-way video feed, and that the judges thought that participants were unaware of their presence. This experiment thus tests how embarrassment is affected by different knowledge levels (shared knowledge vs. common knowledge) and eye contact (eye contact vs. no eye contact) in a 2 × 2 full factorial between-subjects design.

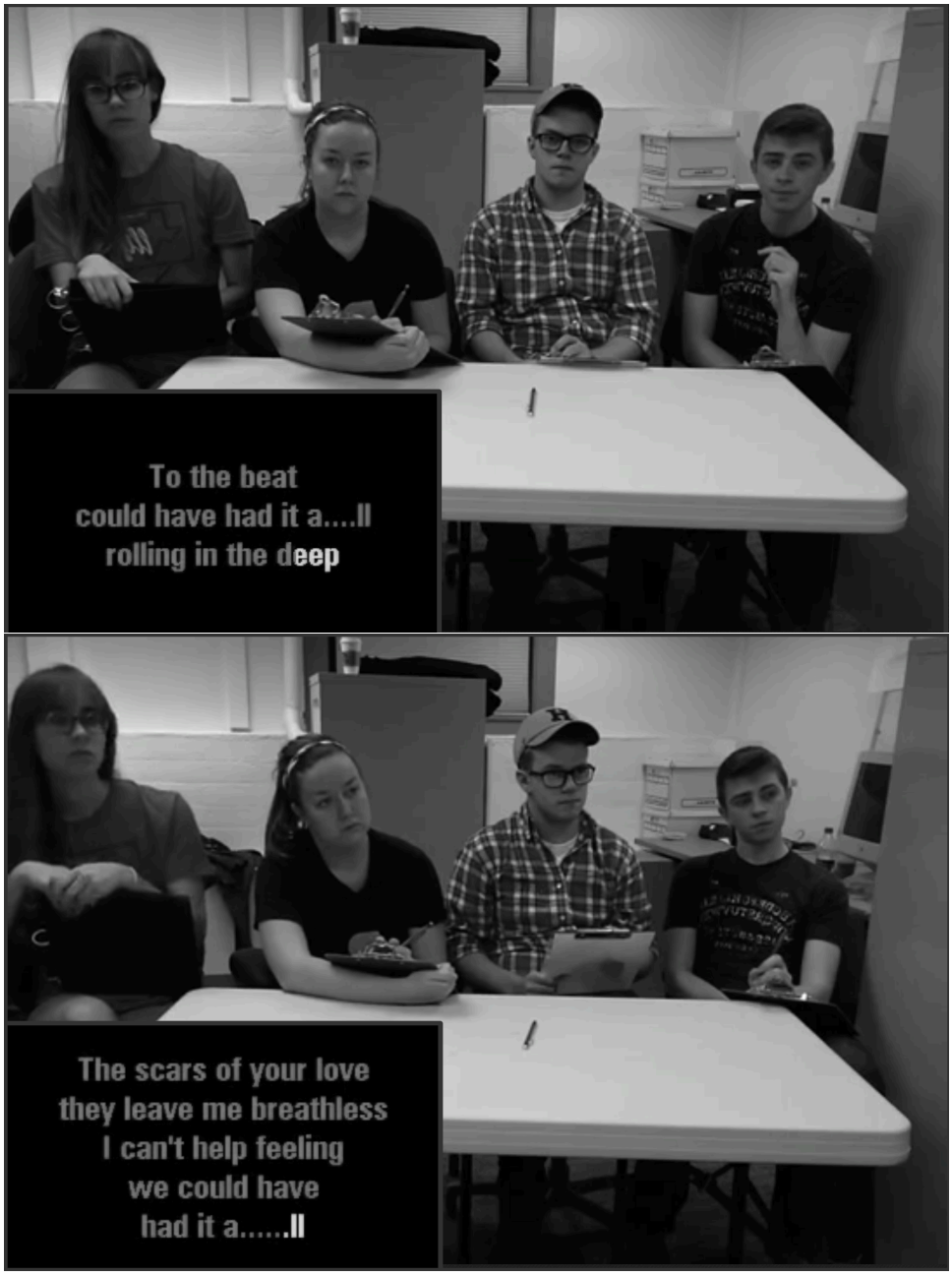

Figure 3.3 Representative screenshots of the videos participants saw in the eye contact (3a, on the top), and no eye contact conditions (3b, on the bottom) in Experiment 2. The black screen with words in the lower left corner was an embedded video that provided lyrics to the karaoke song.

Method

Participants. One hundred and twenty-six participants were recruited from the Harvard University study pool, and paid $10, or given course credit for participation. Thirteen

participants were excluded from analyses, because they expressed suspicion during debriefing (see Procedure), or didn't actually perform the karaoke song, leaving a final sample of $N = 113$ (57% female, $M_{age} = 28.1$, $SD_{age} = 13.3$).

Procedure. Participants signed up for a study described as an experiment on how anonymity affects judges' evaluations of someone's performance, and which stated that they would either be assigned to perform a karaoke song, or to judge someone else's performance. Upon arrival, participants were told that we were studying how the relationship between judges and performers affects judges' evaluations—an important unanswered question in the age of reality TV—and that in each study we recruit a panel of judges and a performer. All participants were then informed that they had been assigned to the performer role, and that four other participants assigned to the role of judges were already situated in a room across the hallway. Participants were then told that they would perform a karaoke song that the judges would observe over a live two-way video feed, and that this two-room set up with the video feed was necessary to ensure that the environment was the same when judges believed they were anonymous and when they did not.

Participants assigned to the *common knowledge condition* were then told: "You have been selected for the condition in which the judges know that you are aware of their presence. You will be able to see them on the two-way video feed during your performance, and they know that you can see them through the two-way video feed. In other words, both you and the judges know that you are watching each other over a two-way video feed during your performance." Participants assigned to the *shared knowledge condition* were instead told: "You have been selected for the anonymous condition, in which the judges think that you are unaware of their presence. However, in order to keep the performances the same across conditions, we actually

always tell the performers about the judges. In addition, you will be able to see them on the two-way video feed during your performance, but they don't know this. In other words, even though you can see the judges over this feed, they think that you don't know they are there, and that they are simply watching a one-way video feed of your performance." Participants in both conditions were then told that after the song ended, the video feed would cut off, and they would be given a short survey[13].

Participants were then escorted into a small private room with a large computer monitor that had a built in video camera, speakers, and a microphone for them to sing into. The researcher set the computer up in front of the participant, and then told the participant that she had to go set up the video feed in the judges' room. Participants were told that as soon as the researcher got this set up, they would see a video feed of the judges as well as a karaoke-style video with the song lyrics, and were instructed to begin their performance as soon as the song started. After 30 seconds, a video appeared on the computer screen with the researcher's head right in front of the camera (as though she was just turning on the video camera in the other room). Immediately thereafter, the song (*Rolling in the Deep* by Adele) began to play, and participants saw a large video showing the judges in the other room, with a small karaoke-style video embedded in the lower left hand corner that displayed the lyrics to the song (see Figure 3.3). In the *eye contact* condition, the video of the judges showed them looking directly into the camera (Figure 3.3a); in the *no eye contact* condition, the video of the judges showed them looking askance, as though they were watching the participant's performance on a screen a few feet to the side of the camera filming them (Figure 3.3b).

[13] Complete instructions provided to participants are supplied in Appendix C of Supplementary Materials

After the song finished, the video feed shut off, and a survey popped up on the computer screen, in which participants were asked to rate how strongly they experienced different emotions on a 1-5 scale in a slightly modified Positive and Negative Affect Schedule (PANAS). The PANAS includes 10 items comprising a negative affect scale, and 10 items comprising a positive affect scale (see Watson, Clark & Tellegen, 1988), and was used to disguise our primary interest in embarrassment, by embedding it in a larger battery of emotion ratings. However, the PANAS does not specifically ask about embarrassment, and so one of the positive affect items was replaced with *embarrassed*.

Participants then answered a few questions about their experience with karaoke, and provided some basic demographic information. Participants then exited the room, where they met the researcher, and were given a funnel debriefing. The funnel debriefing began with very vague questions about what participants thought the study was about, which were followed by more and more specific questions about whether they thought everything had been clearly explained to them, whether they were at all suspicious during the study, and concluded with a direct question about whether they believed the judges were real. Participants were then fully debriefed and paid.

Results

Figure 3.4 displays participants' reported embarrassment in the two conditions, and results showed that participants were more embarrassed in the common knowledge condition. A 2 (shared knowledge vs. common knowledge) × 2 (eye contact vs. no eye contact) ANOVA revealed a marginally significant effect of knowledge level on ratings of embarrassment, $F(3, 109) = 3.83$, $p = .053$, $\eta_p^2 = .03$ (see Figure 3.4a), but no significant effect of eye contact, $F(3, 109) = 0.96$, $p = .328$, $\eta_p^2 = .01$, or significant interaction, $F(3, 109) = 0.15$, $p = .703$, $\eta_p^2 = .00$.

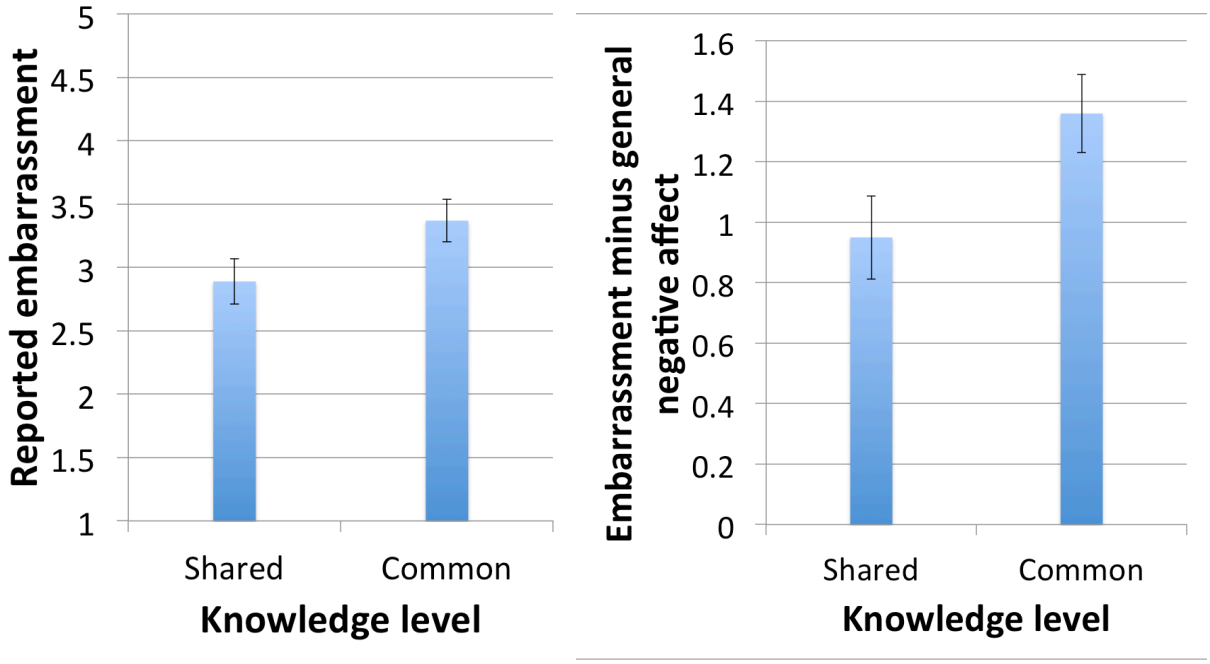

Figure 3.4 Average reported embarrassment (3.4a, on the left), and average reported embarrassment with general negative affect subtracted out (3.4b, on the right), in the shared and common knowledge conditions in Experiment 2. Error bars represent standard error.

However, as the results of Experiment 1 suggest, the effects of shared versus common knowledge can be relatively specific to the self-conscious emotions. Hence, we wanted to separate participants' embarrassment ratings from their general negative affect. We estimated negative affect by averaging the 10 items of the negative affect scale from the PANAS ($\alpha = .891$) and then we subtracted this value from ratings of embarrassment. A 2 × 2 ANOVA revealed a slightly stronger effect of knowledge level on this refined measure of embarrassment, $F(3, 109) = 4.77$, $p = .031$, $\eta_p^2 = .04$ (see Figure 3.4b), but no significant effect of eye contact, $F(3, 109) = 1.82$, $p = .181$, $\eta_p^2 = .02$, or significant interaction, $F(3, 109) = 0.16$, $p = .691$, $\eta_p^2 = .00$. These results support the prediction that participants would be more embarrassed when their performance was common knowledge than shared knowledge, but provide no support for the prediction that eye contact would increase embarrassment.

Discussion

This study provides evidence from two experiments that a transgression or faux pas evokes more intense negative self-conscious emotions when it is common knowledge than when it is shared knowledge. In Experiment 1, participants reported that they would feel more embarrassed, guilty, and ashamed in scenarios when an offense was common versus shared knowledge. In Experiment 2, participants reported feeling more embarrassed by a karaoke performance in front of a panel of judges when they believed it was common knowledge than when they believed it was shared knowledge. Previous research has shown that people experience more intense self-conscious emotions when an audience is present. These results suggest that the level of knowledge an individual shares with the audience also matters: Realizing that other people know about one's offense can be uncomfortable, but it is even more agonizing if the offense becomes common knowledge.

Why do the self-conscious emotions function this way? In the introduction we suggested this could serve a strategic function, because there are situations in which an individual's best move depends on whether something is shared or common knowledge. These types of situations are modeled in game theory as *coordination games*, and both relationships and social status can be characterized as coordination games. In a coordination game, individuals can benefit by coordinating their behavior, but there is more than one way to do so, and coordination failures are costly. For example, everyone can benefit by adhering to established social norms, but these can take many different forms, and failing to coordinate on the same norms can be costly (e.g., a foreigner who fails to recognize some local custom may face condemnation or social exclusion). Likewise, people can benefit by coordinating on one of many possible types of relationships (e.g., friends, lovers, boss-employee, etc.), each of which is governed by distinct expectations and norms (Fiske, 1992; Haslam, 2004; Lee & Pinker, 2010), and coordination failures can be

costly (e.g., getting fired for behaving inappropriately in front of a boss when one gets too friendly). Social status can also be modeled as a coordination game, because it is an abstract social construction based solely on the prestige, power, and deference collectively granted to someone by others. Here too, coordination failures can be costly, as when one invites costly retaliation by failing to acknowledge the high status of a powerful individual, or squanders effort and social capital by sucking up to someone with low status. We suggest that two core functions of the self-conscious emotions—managing relationships and social status—explain the sensitivity of self-conscious emotions to these different knowledge states, because both involve the strategic logic of coordination.

If the self-conscious emotions evolved according to such strategic logic, this might help explain another one of their distinctive features: They involve not only facial displays like basic emotions but also dynamic, full body postures (Keltner, 1995; Keltner & Buswell, 1997; Lewis, 2000; Tracy & Matsumoto, 2008; Tracy & Robins, 2004). Shame and embarrassment are associated with slumping the shoulders, hanging the head, and touching the face. Likewise, much of the discomfort of blushing seems to come from the fact that the blusher knows onlookers know they know they are blushing, and so on (see Thomas et al., 2014). Pride, a positive self-conscious emotion, is associated with puffing the chest and raising the arms. These full-bodied, dynamic displays might function to manipulate knowledge levels among senders and receivers. In short, people's evolved strategies and recursive mental state reasoning might explain why some of the key characteristics of self-conscious emotion signals differ from those of the basic emotions.

Eye contact did not lead to greater embarrassment in Experiment 2. It could be the case that eye contact doesn't provide a sufficient cue to common knowledge, contra our hypothesis.

Or, this null result may have been due to a relatively weak manipulation, because eye contact was video-mediated, and even in the no eye contact condition participants could still see judges' eyes (see Figure 3.3b). Given the prevalence and importance of eye contact in human social life (see for example, Baron-Cohen, 1997), we suspect our experimental design simply failed to capture the phenomenon sufficiently, and believe further research is warranted before drawing any strong conclusions.

One unexpected result was that participants reported greater anger in the common knowledge condition than in the shared knowledge condition. The two basic emotion scenarios were included to test whether the different effects of shared and common knowledge are specific to self-conscious emotions. As explained above, the hypothesized effect of shared versus common knowledge on the self-conscious emotions was based on these emotions' fundamental role in managing relationships and social status, and the coordination dynamics involved in both endeavors. However, basic emotions such as anger and sadness can also serve such social functions (Frank, 1988; Reed, DeScioli, & Pinker, 2014; Sell, Tooby, & Cosmides, 2009), and may thus be sensitive to the distinction between shared and common knowledge in some social circumstances.

We close by noting that these results bolster previous findings showing that the categorical difference between shared and common knowledge—long recognized by game theorists (Rubinstein, 1989; Schelling, 1960), economists (Chwe, 2001; Geanakoplos, 1992), philosophers (Lewis, 1969), linguists (Clark, 1996; Clark & Marshall, 1981; Smith, 1982), sociologists (Adut, 2008), anthropologists and political scientists (Cronk & Leech, 2012), legal theorists (McAdams & Nadler, 2005), and even computer scientists (Alberucci & Jäger, 2005; Halpern & Moses, 1990)—has important and far-reaching psychological implications. Previous

research has demonstrated that the distinction between shared and common knowledge matters for cooperation and coordination (Thomas et al., 2014), and plays an important role in explaining strategic uses of indirect speech (Lee & Pinker, 2010). This study adds self-conscious emotions to the list of human psychological systems that are sensitive to the difference between shared and common knowledge.

CONCLUSION

These studies provide the first direct evidence for a psychology of common knowledge. They also hint at its potentially wide-ranging applications, by demonstrating how common knowledge can affect arenas of human social life in which important underlying aspects of coordination are implicit and unobvious. For example, the effects of common knowledge on indirect speech and self-conscious emotions that are explored in Papers 2 and 3 seem to be much less obvious than the more explicit decision of whether or not to try and work together in the coordination games employed in Paper 1. The depth and breadth of theory on the topic suggests both that such empirical work is long overdue, and that the findings presented here merely scratch the surface of a potentially vast unexplored area of human psychology and social behavior.

The first paper shows that people recognize common knowledge, differentiate it from lower levels of shared knowledge, and utilize these different knowledge states strategically. In the experiments reported in the paper, participants engaged in interactive scenarios that embodied the coordination game Rubinstein (1989) used in his famous proof. These studies were designed to test how different levels of knowledge would affect behavior in a canonical coordination game. The results unequivocally supported the hypothesis that common knowledge leads people to attempt to coordinate with others for mutual benefit, even when doing so is risky and entails a large opportunity cost. These findings provide an important proof of concept, by showing that people recognize the strategic importance of different kinds of knowledge when trying to coordinate with others, and that common knowledge plays an important role in facilitating coordination.

These studies also offer multiple converging lines of evidence suggesting that the mind has a dedicated representation for common knowledge, which is categorically distinct from representations for lower levels of shared knowledge. Distinct representations for these two kinds of knowledge could provide evidence of adaptive design, because this mirrors the categorical difference between these knowledge states in game theory models of coordination; however, further research is needed to confirm this conclusion.

The results reported in the second paper support a strategic theory of indirect speech, which postulates that indirect speech offers speakers an optimal way to try and transition between relationships with a listener, by preventing common knowledge of a proposition. In the study, participants engaged in two scripted interactions, one involving a bribe proposition, the other a sexual proposition, while cardiovascular measures were taken to assess their affective responses. Participants' responses to direct and indirect propositions in the two scenarios were consistent with the social and reputational costs described by the theory. Participants' responses to survey questions about the sexual proposition also supported the theory's prediction that indirect speech can help speakers prevent rumors from spreading. Even though this study did not directly test any of the theory's predictions regarding common knowledge, by supporting the overall theory, these results provide indirect evidence that common knowledge plays an important role in explaining strategic uses of indirect speech. It is worth noting that in other studies not reported here, we have found more direct evidence for the role of common knowledge in explaining indirect speech.

The third paper provides evidence from two studies that the self-conscious emotions of shame, guilt, and embarrassment are sensitive to the distinction between

shared and common knowledge. In the first study, participants read fictional scenarios in which they committed some transgression, and reported that they would expect to feel more ashamed, embarrassed, and guilty when their transgression was common knowledge than when it was shared knowledge. A similar pattern of results was observed in a second lab study, in which participants performed a karaoke song for a panel of judges over a two-way video feed. In this study, participants who believed the two-way video feed was commonly known reported higher levels of embarrassment, than participants who believed the judges erroneously believed they were watching a one-way video feed of the participant's performance, and that the participant was unaware of their presence (i.e., the performance was shared knowledge). This paper thus shows another psychological context in which common knowledge plays a role, and that it affects the self-conscious emotions of embarrassment, shame, and guilt.

Taken together, these three papers suggest that our evolved psychology has been shaped by the strategic demands of coordination, and that the mind contains computational mechanisms designed to recognize common knowledge. Furthermore, they demonstrate that common knowledge applies to more than just situations involving the kinds of explicit coordination problems that are typically modeled by game theorists (e.g., situations in which people must decide whether or not to work together for mutual benefit, such as those employed in Paper 1). Throughout this dissertation I have argued that common knowledge likely plays a role in many arenas of human social life, yet I have only provided evidence for a few. However, by demonstrating that common knowledge is more than just a theoretical curiosity, I believe this opens the door for understanding its many other diverse applications.

In closing, I would like to suggest an analogy that I believe helps clarify why common knowledge might matter in so many different contexts. I have suggested that the one thing that ties all these contexts together is that they involve coordination, which is backed by research in game theory. Coordination games are an abstract mathematical concept (like perfect circles or normal distributions) that merely provide a structural account of a situation, but say nothing about the possible semantic contexts that might embody such a structure. I would like to suggest that common knowledge is a deep mathematical concept (much like π or eigenvalues), which pops up all over the place because of deep underlying isomorphisms in contexts that can on the surface seem completely unrelated. Just as π pops up in explanations of many physical phenomena that share deep underlying geometric isomorphisms, and eigenvalues pop up in many different informational contexts because of deep underlying isomorphisms in information structures, I suggest that common knowledge pops up all over the place in human social life because of deep underlying isomorphisms involving coordination. Now it's time for the necessary empirical work to be done to elaborate and confirm this conjecture.

BIBLIOGRAPHY

Abele, S., Stasser, G., & Chartier, C. (2013). Use of social knowledge in tacit coordination: Social focal points. *Organizational Behavior and Human Decision Processes, 123*(1), 23-33. doi:10.1016/j.obhdp.2013.10.005

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology, 19*(6), 586–592. doi:10.1037//0278-6133.19.6.586

Adut, A. (2008). *On scandal: Moral disturbances in society, politics, and art*. New York, NY: Cambridge University Press.

Alberucci, L., & Jäger, G. (2005). About cut elimination for logics of common knowledge. *Annals of Pure and Applied Logic, 133*(1-3), 73-99. doi:10.1016/j.apal.2004.10.004

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the Internet: The effect of $1 stakes. *PLoS ONE, 7*(2), e31461. doi:10.1371/journal.pone.0031461

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953-970. doi:10.1037/a0016923

Argyle, M., & Furnham, A. (1983). Sources of satisfaction and conflict in long-term relationships. *Journal of Marriage and Family, 45*(3), 481-493. Retrieved from http://www.jstor.org/stable/351654

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics, 4*(6), 1236-1239. doi:10.1214/aos/1176343654

Austin, J. L. (1962). *How to do things with words, 2nd ed.* J. O. Urmson & M. Sbisa (Eds.). Cambridge, MA: Harvard University Press.

Ayres, I., & Nalebuff, B. J. (1996). Common knowledge as a barrier to negotiation. *UCLA Law Review, 44*, 1631-1659.

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2008, December). *Explaining focal points: Cognitive hierarchy theory versus team reasoning.* Discussion paper, Centre for Decision Research and Experimental Economics. Retrieved from http://www.nottingham.ac.uk/cedex/documents/papers/2008-17.pdf

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: The MIT Press.

Barry, E. (2011, July 14). Sound of post-Soviet protest: Claps and beeps. *The New York Times*. Retrieved from http://www.nytimes.com/2011/07/15/world/Europe/15belarus.html?pagewanted=all

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments is a fundamental human motivation. *Psychological Bulletin, 117*(3), 497-529. doi:10.1037/0033-2909.117.3.497

Becker, G. S. (1976). *The economic approach to human behavior.* Chicago, IL: The University of Chicago Press.

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., Knight, R. T. (2003). The regulatory function of self-conscious emotions: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology, 85*(4), 594-604. doi:10.1037/0022-3514.85.4.594

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2013). Separating the shirkers from the workers? Making sure respondents pay attention to self-administered surveys. *American Journal of Political Science.* doi:10.1111/ajps.12081

Bernston, G. G., Quigley, K. S., & Lozano, D. (2007). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 182-210). New York, NY: Cambridge University Press.

Binmore, K., & Samuelson, L. (2001). Coordinated action in the electronic mail game. *Games and Economic Behavior, 35,* 6-30. doi:10.1006/game.1998.0749

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14*(3), 283-286. doi:10.1111/1467-9280.03436

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*(5), 382-386. doi:10.1111/j.1467-9280.2007.01909

Blascovich, J., & Mendes, W. B. (2010). Social psychophysiology and embodiment. In S. T. Fiske, D. T. Gilbert, & Lindzey, G. (Eds.), *The handbook of social psychology* (5th ed., pp. 194-227). New York, NY: John Wiley & Sons Inc.

Blascovich, J., Mendes, W. B., Tomaka, J., Salomon, K., & Seery, M. (2003). The robust nature of the biopsychosocial model challenge and threat: A reply to Wright and Kirby. *Personality and Social Psychology Review, 7*(3), 234-243. doi:10.1207/S15327957PSPR0703_03

Bosch, J. A., de Gues, E. J. C., Carroll, D., Goedhart, A. D., Anane, L. A., Veldhuizen, J. J., … Edwards, K. M. (2009). A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosomatic Medicine, 71*, 877-885. doi:0.1097/PSY.0b013e3181baef05

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York, NY: Cambridge University press.

Brownley, K. A., Hurwitz, B. E., & Schneiderman, N. (2000). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 224-264). New York, NY: Cambridge University Press.

Buchan, N. R., Croson, R. T. A., & Dawes, R. M. (2002). Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange. *American Journal of Sociology, 108*(1), 168-206. doi:10.1086/344546

Buss, D. M. (1994). *The evolution of desire*. New York, NY: Basic Books.

Byman, D. (2013, May 4). Mr. Obama, don't draw that line. *The New York Times*. Retrieved from http://www.nytimes.com/2013/05/05/opinion/sunday/dont-draw-that-red-line.html?_r=0&adxnnl=1&pagewanted=all&adxnnlx=1381876104-nShk8Dh8i2A3tkeL5o+IJA

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., Tapanya, S., & Singh, S. (2005). Synchrony in the onset of mental-state reasoning. *Psychological Science, 16*(5), 378-384. doi:10.1111/j.0956-7976.2005.01544.x

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction.* Princeton, NJ: Princeton University Press.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1-3), 7-42. doi:10.1023/A:1007850605129

Camerer, C. F., Loewenstien, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97*(5), 1232-1254. Retrieved from http://www.jstor.org/stable/1831894

Centola, D., Willer, R., & Macy, M. (2005). The emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology, 110*(4), 1009-1040. doi:10.1086/427321

Chakroff, A., Thomas, K. A., Haque, O. S., & Young, L. (2014). An indecent proposal: The dual functions of indirect speech. *Cognitive Science*, 1-13. doi:10.1111/cogs.12145

Chartier, C., Abele, S., Stasser, G., & Shriver, E. (2012, January). *Secondary and Schelling salience in matching and mismatching coordination*. Poster session presented at the 13th annual Society for Personality and Social Psychology meeting, San Diego, CA.

Chaudhuri, A., Schotter, A., Sopher, B. (2009). Talking ourselves to efficiency: Coordination in

inter-generational minimum effort games with private, almost common and common knowledge of advice. *The Economic Journal, 119*(534), 91-122. doi:10.1111/j.1468 0297.2008.02207.x

Chwe, M. S. (2001). *Rational ritual: Culture, coordination, and common knowledge*. Princeton, NJ: Princeton University Press.

Clark, H. H. (1992). *Arenas of language use.* Chicago, IL: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, UK: Cambridge University Press.

Clark, R. D., & Hatfield, E. (1989). Gender differences in receptivity to sexual offers. *Journal of Psychology and Human Sexuality, 2*(1), 39-55. doi:10.1300/J056v02n01_04

Clifford, S., & Jerit, J. (2012). Do attempts to improve respondent attention increase social desirability bias? In *APSA 2012 Annual Meeting Paper.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2105228

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences, 26*(2), 139-198. doi:10.1017/S0140525X03000050

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences, 106*(23), 9163-9168. doi:10.1073/pnas.0807721106

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). Oxford, UK: Oxford University Press.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 584-627). Hoboken, NJ: Wiley.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1)*,* 87-185. doi:10.1017/S0140525X01003922

Cronk, L. & Leech, B. L. (2013). *Meeting at Grand Central: Understanding the social and evolutionary roots of cooperation.* Princeton, NJ: Princeton University Press.

Curry, O., & Chesters, M. J. (2012). 'Putting ourselves in the other fellow's shoes': The role of

'Theory of Mind' in solving coordination problems. *Journal of Cognition and Culture, 12*, 147-159. doi:10.1163/156853712X633974

Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science, 17*(12), 1082-1089. doi:10.1111/j.1467-9280.2006.01834.x

Dalkiran, N. A., Hoffman, M., Paturi, R., Ricketts, D., & Vattani, A. (2012). Common knowledge and state-dependent equilibria. In *Proceedings of Symposium on Algorithmic Game Theory* (pp. 84-95). doi:10.1007/978-3-642-33996-7_8

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs, & N. B. Davies (Eds.), *Behavioral ecology: An evolutionary approach* (pp. 282-309). Oxford, UK: Blackwell.

DeScioli, P., Bruening, R., & Kurzban. R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior*, *32*, 204-215. doi:10.1016/j.evolhumbehav.2011.01.003

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*, 477-496. doi:10.1037/a0029065

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*(5), 630-641. doi:10.1037//0012-1649.37.5.630

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010, April). Are your participants gaming the system? Screening Mechanical Turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2399-2402). ACM. doi:10.1145/1753326.1753688

Ekman, P. (2003). *Emotions revealed*. New York, NY: Times Books.

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: Reactive egoism in social interaction. *Journal of Personality and Social Psychology, 91*(5), 872-889. doi:10.1037/0022-3514.91.5.872

Feiler, B. (2000, October). Pocketful of dough. *Gourmet,* 99-101. Retrieved from http://www.gourmet.com/magazine/2000s/2000/10/pocketful

Fischer, I. (2009). Friend or foe: Subjective expected relative similarity as a determinant of cooperation. *Journal of Experimental Psychology: General, 138*(3), 341-350. doi:10.1037/a0016073

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review, 99*(4)*,* 689-723. doi:10.1037/0033-295X.99.4.689

Fiske, A. P. (2004). Four modes of constituting relationships: Consubstantial assimilation; space, magnitude, time, and force; concrete procedures; abstract symbolism. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 61-146). Mahwah, NJ: Erlbaum.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York, NY: W. W. Norton.

Friedell, M. (1967). On the structure of shared awareness. *Working papers of the Center for Research on Social Organizations*. Ann Arbor: University of Michigan, paper 27. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/50812

Friedman, M. (1991). The island of stone money (No. e-91-3). Retrieved from http://hoohila.stanford.edu/workingpapers/getWorkingPaper.php?filename=E-91-3.pdf

Frith, U., & Frith, C. D. (2003). Developmental and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B, 358*, 459-473. doi:10.1098/rstb.2002.1218

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501. doi:10.1016/S1364-6613(98)01262-5

Geanakoplos, J. (1992). Common knowledge. *The Journal of Economic Perspectives, 6*(4), 53-82. doi:10.1257/jep.6.4.53

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* Oxford, UK: Oxford University Press.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 89-150). New York, NY: McGraw-Hill.

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry*, *18*(3), 213-231. doi:10.1521/00332747.1955.11023008

Goffman, E. (1959). *The presentation of self in everyday life.* New York, NY: Anchor Books.

Goldberg, L. R. (1992). The development of markers of the Big Five factor structure. *Psychological Assessment, 4,* 26-42. doi:10.1037/1040-3590.4.1.26

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Dreary, F. De Fruyt, & F. Ostendord (Eds.), *Personality psychology in Europe*, vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2012). Data collection in a flat world: The

strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*(3), 213-224. doi:10.1002/bdm.1753

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology, 93*(4), 583-599. doi:10.1037/0022-3514.93.4.583

Graziano, W. G., & Tobin, R. M. (2002). Agreeableness: Dimension of personality or social desirability artifact? *Journal of Personality, 70*(5), 695-728. doi:10.1111/1467-6494.05021

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax & semantics 3: Speech acts* (pp. 41-58). New York, NY: Academic Press.

Gruenewald, T. L., Kemeny, M. E., Aziz, N., & Fahey, J. L. (2004). Acute threat to the social self: Shame, social self-esteem, and cortisol activity. *Psychosomatic Medicine, 66*, 915-924. doi:10.1097/01.psy.0000143639.61693.ef

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26*, 245-256. doi:10.1016/j.evoljumbehav.2005.01.002

Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM, 37*(3), 549-587. doi: 10.1145/79147.79161

Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others' goals. *Developmental Science, 11*(4), 487-494. doi:10.1111/j.1467-7687.2008.00694.x

Haslam, N. (2004). Research on the relational models: An overview. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 27-57). Mahwah, NJ: Erlbaum.

Helgeson, V. S., Shaver, P., & Dyer, M. (1987). Prototypes of intimacy and distance in same-sex and opposite-sex relationships. *Journal of Social and Personal Relationships, 4*(2), 195-233. doi:10.1177/0265407587042006

Hoffman, M., & Yoeli, E. (2013, September 14). Effective red lines. *The Economist*. Retrieved from http://www.economist.com/news/letters/21586273-syria-chemical-weapons-high-speed-rail-azerbaijan-italy-burials

Hofstadter, D. (1985). *Metamagical themas: Questing for the essence of mind and pattern.* New York, NY: Basic Books.

Holtgraves, T. M., & Yang, J. N. (1990). Politeness as universal: Cross-cultural perceptions of

request strategies and inferences based on their use. *Journal of Personality and Social Psychology*, *59*(4), 719-729. doi:10.1037/0022-3514.59.4.719

Horn, L. R. (2004). Implicature. In L. R. Horn & G. Ward (Eds.) *The handbook of pragmatics* (pp. 3-28). Malden, MA: Blackwell Publishing Ltd.

Horton, J. J, Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*(3), 399-425. doi:10.1007/s10683-011-9273-9

Hume, D. (1969). *A treatise of human nature*. London, UK: Penguin. (Original work published 1739-1740).

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*(3), 441-454. doi:10.1037/0022-3514.68.3.441

Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion, 10*(2), 155-171. doi:10.1080/026999396380312

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*(3), 250-270. doi:10.1037/0033-2909.122.3.250

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25-41. doi:10.1016/S0010-0277(03)00064-7

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology, 89,* 191-204. doi:10.1111/j.2044 8295.1998.tb02680.x

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology, 75*(5), 1155-1178. doi:10.1037/0022-3514.75.5.1155

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, IL: University of Chicago Press.

Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences, 31*(2), 215-226. doi:10.1016/S0191-8869(00)00130-6

Lee, J. J., Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review, 117*(3), 785-807. doi:10.1037/a0019688

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University

Press.

Lewis, M. (2000). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 623–636). New York, NY: Guilford.

Li, N. P., & Kenrick, D. T. (2006). Sex similarities and differences in preferences for short-term mates: What, whether, and why. *Journal of Personality and Social Psychology, 90*(3), 468-489. doi:10.1037/0022-3514.90.3.468

Lieberman, D., Oum, R., & Kurzban, R. (2008). The family of fundamental social categories includes kinship: Evidence from the memory confusion paradigm. *European Journal of Social Psychology, 38*(6), 998-1012. doi:10.1002/ejsp.528

Lozano, D. L., Norman, G., Knox, D., Wood, B. L., Miller, B. D., Emery, C. F., & Bernston, G. G. (2007). Where to B in dZ/dt. *Psychophysiology, 44*(1), 113-119. doi:10.1111/j.1469-8986.2006.00468.x

Marr, D. (1982). *Vision*: *A computational investigation into the human representation and processing of information*. New York, NY: W. H. Freeman.

Maynard-Smith, J., & Harper, D. (2003). *Animal signals.* Oxford, UK: Oxford University Press.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*(03), 353-363. Retrieved from http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=29989&fileId=S0140525X98001228

McAdams, R. H., & Nadler, J. (2005). Testing the focal point theory of legal compliance: The effect of third-party expression in an experimental hawk/dove game. *Journal of Empirical Legal Studies, 2*(1), 87-123. doi:10.1111/j.1740-1461.2005.00032.x

McCullough, M. E. (2008). *Beyond revenge: The evolution of the forgiveness instinct.* San Francisco, CA: Jossey-Bass.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-444. doi:10.1146/annurev.soc.27.1.415

Mehta, J., Starmer, C., & Sugden, R. (1994a). Focal points in pure coordination games: An experimental investigation. *Theory and Decision, 36*(2), 163-185. doi:10.1007/BF01079211

Mehta, J., Starmer, C., & Sugden, R. (1994b). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review, 84*(3), 658-673. Retrieved from http://www.jstor.org/discover/10.2307/2118074?uid=2&uid=4&sid=21106126413391

Mendes, W. B. (2009). Assessing the autonomic nervous system. In E. Harmon-Jones, & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 118-147). New York, NY: Guilford Press.

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: Physiological responses during social interactions with expectancy-violating partners. *Journal of Personality and Social Psychology, 92*(4), 698-716. doi:10.1037/0022-3514.92.4.698

Mendes, W. B., Major, B., McCoy, S., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *Journal of Personality and Social Psychology, 94*(2), 278-291. doi:10.1037/0022-3514.94.2.278

Mendes, W. B., & Park, J. (2014). Neurobiological concomitants of motivational states. In A. J. Elliot (Ed.), *Advances in motivation science* (1st ed.) (pp. 233-270). Waltham, MA: Elsevier Inc.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4*, 1-25. doi:10.1016/0022-1031(68)90046-2

Milinski, M., Semmann, D., Krambeck, H. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature, 415*, 424-426. doi:10.1038/415424a

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America, 27*(2), 338-352. doi:10.1121/1.1907526

Miller, R. S. (1995). On the nature of embarrassability: Shyness, social evaluation, and social skill. *Journal of Personality, 63*(2), 315-339. doi:10.1111/j.1467-6494.1995.tb00812.x

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50*(4), 655-663. doi:10.1016/j.neuron.2006.03.040

Modigliani, A. (1971). Embarrassment, facework, and eye contact: Testing a theory of embarrassment. *Journal of Personality and Social Psychology, 17*(1), 15-24. doi:10.1037/h0030460

Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior, 1*(2), 170-190. doi:10.1016/0899-825689900171

Morris, S. (2002). Coordination, communication, and common knowledge: A retrospective on the electronic-mail game. *Oxford review of economic policy, 18*(4), 433-445. doi:10.1093/oxrep/18.4.433

Nicholson, N., Soane, E., Fenton-O'Creevy, M., & Willman, P. (2005). Personality and domain specific risk taking. *Journal of Risk Research, 8*(2), 157-176. doi:10.1080/1366987032000123856

Nov, O., & Rafaeli, S. (2009). Measuring the premium on common knowledge in computer mediated coordination problems. *Computers in Human Behavior, 25*(1), 171-174. doi:10.1016/j.chb.2008.08.005

Ockenfels, A., & Werner, P. (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior and Organization*. doi:10.1016/j.jebo.2013.12.003

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258. doi:10.1126/science.1107621

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872. doi:10.1016/j.jesp.2009.03.009

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge, UK: Cambridge University Press.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411-419. Retrieved from http://journal.sjdm.org/10/10630a/jdm10630a.html

Pedersen, E. J., Kurzban, R., & McCullough, M. E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B: Biological Sciences, 280*(1758). doi:10.1098/rspb.2012.2723

Perner, J., & Wimmer, H. (1985). "John *thinks* that Mary *thinks* that…" attribution of second order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437-471. doi:10.1016/0022-0965(85)90051-7

Pierce, J. R., Kilduff, G. J., Galinsky, A. D., & Sivanathan, N. (2013). From glue to gasoline: How competition turns perspective takers unethical. *Psychological Science, 24*(10), 1986-1994. doi:10.1177/0956797613482144

Pinker, S. (2000). Survival of the clearest. *Nature, 404*, 441-442. doi:10.1038/35006523

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York, NY: Viking.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences, 105*(3), 833-838. doi:10.1073/pnas.0707192105

Provine, R. R. (1996). Laughter. *American Scientist, 84*(1), 38-45. Retrieved from http://www.jstor.org/stable/29775596

Provine, R. R. (2012). *Curious behavior: Yawning, laughing, hiccupping, and beyond.* Belnap Press.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*, 427-430. doi:10.1038/nature11467

Reed, L. I., DeScioli, P., & Pinker, S. (2014). The commitment function of angry facial expressions. *Psychological Science, 25*(8), 1511-1517. doi:10.1177/0956797614531027

Richardson, D. S. (2014). Everyday aggression takes many forms. *Current Directions in Psychological Science, 23*(3), 220-224. doi:10.1177/0963721414530143

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin, 28*(6), 789-801. doi:10.1177/0146167202289008

Rousseau, J. (1984). *A discourse on inequality.* (M. Cranston, Trans.). New York, NY: Penguin Books. (Original work published 1754).

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge." *The American Economic Review, 79*(3), 385-391. Retrieved from http://www.jstor.org/stable/1806851

Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The Investment Model Scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships, 5*, 357-391. doi:10.1111/j.1475-6811.1998.tb00177.x

Saxe, R., & Young, L. (2013). Theory of mind: How brains think about thoughts. In K. Ochsner & S. Kosslyn (Eds.), *The Handbook of Cognitive Neuroscience*. Oxford, UK: Oxford University Press.

Scaife, M, & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*(5489), 265-266. doi:10.1038/253265a0

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Searle, J. (1989). How performatives work. *Linguistics and Philosophy, 12*(5), 535-558. doi:10.1007/BF00627773

Searle, J. (1995). *The construction of social reality*. New York, NY: The Free Press.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*(35), 15073-15078. doi:10.1073pnas.0904312106

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & van Dooren, L. J. P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology, 27*, 1-23. doi:10.1111/j.1469-8986.1990.tb02171.x

Skyrms, B. (2004). *The stag hunt and the evolution of social structure.* Cambridge, UK: Cambridge University Press.

Smith, N. (1982). *Mutual knowledge.* Orlando, FL: Academic Press.

Smith, R. H., Webster, J. M., Parrot, W. G., & Eyre, H. L. (2002). The role of public exposure in moral and nonmoral shame and guilt. *Journal of Personality and Social Psychology, 83*(1), 138-159. doi:10.1037//0022-3514.83.1.138

Steckler, C. M., & Tracy, J. L. (2014). The emotional underpinnings of social status. In J. T. Cheng, J. L. Tracy, & C. Anderson (Eds.), *The psychology of social status* (pp. 201-224). New York, NY: Springer.

Sugden, R. (1995). A theory of focal points. *The Economic Journal, 105*(430), 533-550. Retrieved from http://www.jstor.org/stable/2235016

Summerville, A., & Chartier, C. R. (2013). Pseudo-dyadic "interaction" on Amazon's Mechanical Turk. *Behavior Research Methods, 45*(1), 116-124. doi:10.3758/s13428-012-0250-9

Sun, T., & Wu, G. (2012). Traits, predictors, and consequences of Facebook self-presentation. *Social Science Computer Review, 30*(4), 419-433. doi:10.1177/0894439311425978

Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE, 6*(3), e16836. doi:10.1371/journal.pone.0016836

Symons, D. (1979). The evolution of human sexuality. New York, NY: Oxford University Press.

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General, 142*(1), 151-162. doi:10.1037/a0028232

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology, 20*(6), 1256-1269. doi:10.1037/0022-3514.70.6.1256

Tangney, J. P., & Tracy, J. L. (2012). Self-conscious emotions. In M. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 446-478). New York, NY: Guilford.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36,* 778-793. doi:10.1037/0022-3514.36.7.778

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology, 107*(4), 657-676. doi:10.1037/a0037037

Tocci, R. J., Widmer, N. S., & Moss, G. L. (2007). *Digital systems: Principles and applications.* Upper Saddle River, NJ: Pearson Education Inc.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. Dunham (Eds.), *Joint attention: Its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum.

Tooby, J., & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 584-627). Hoboken, NJ: Wiley.

Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human morality & sociality: Evolutionary and comparative perspectives* (pp. 91-234). New York, NY: Palgrave MacMillan.

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for *n*-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27*, 103-129. doi:10.1002/mde.1287

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences, 105*(33), 11655-11660. doi:10.1073/pnas.0802686105

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry: An International Journal for the advancement of Psychological Theory, 15*(2), 103-125. doi:10.1207/s15327965pli1502_01

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*(1), 35-57. Retrieved from http://www.jstor.org/stable/2822435

Trivers, R. L. (1985). *Social evolution*. Reading, MA: Cummings.

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*(2), 337-349. doi:10.1037/0022-3514.77.2.337

Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review, 10*(4), 354-371. doi:10.1207/s15327957pspr1004_5

Veblen, T. (1899/2007). *The theory of the leisure class*. Oxford, UK: Oxford University Press.

Walster, E., Berscheid, E., & Walster, G. W. (1973). New directions in equity research. *Journal of Personality and Social Psychology, 25*(2), 151-176.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. doi:10.1037/0022-3514.54.6.1063

Wegner, D. M. (1995). A computer network model of human transactive memory. *Social Cognition, 13*, 1-21. doi:10.1521/soco.1995.13.3.319

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology, 61,* 923-929. doi:10.1037/0022-3514.61.6.923

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684. doi:10.1111/1467-8624.00304

Whiten, A., & Byrne, R. W. (1997). *Machiavellian intelligence II: Extensions and evaluations*. Cambridge, UK: Cambridge University Press.

Willer, R., Kuwabara, K., & Macy, M. (2009). The false enforcement of unpopular norms. *American Journal of Sociology, 115*(2), 451-490. doi:10.1086/599250

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128. doi:10.1016/0010-0277(83)90004-5

Wolf, S. T., Cohen, T. R., Panter, A. T., & Insko, C. A. (2010). Shame proneness and guilt proneness: Toward the further understanding of reactions to public and private transgressions. *Self and Identity, 9*, 337-362. doi:10.1080/15298860903106843

Yamagishi, T., Mifune, N., Liu, J. H., Pauling, J. (2008). Exchanges of group-based favours: Ingroup bias in the prisoner's dilemma game with minimal groups in Japan and New Zealand. *Asian Journal of Social Psychology, 11,* 196-207. doi:10.1111/j.1467-839X.2008.00258.x

Zuckerman, E. W. (2010). What if we had been in charge? The sociologist as builder of rational institutions. In M. Lounsbury & P. M. Hirsch (Eds.) *Markets on trial: The economic sociology of the U.S. financial crisis: Part B (research in the sociology of organizations, volume 30)* (pp.359-378). Emerald Group Publishing Limited. doi:10.1108/S0733-558X(2010)000030B016

SUPPLEMENTARY MATERIALS FOR PAPER 3

Appendix A: Table of Mean Emotion Ratings in Experiment 1

Table SM1 Mean emotion ratings across knowledge levels in Experiment 1

Scenario	Emotion	Knowledge level condition *M*(*SD*)					
		1	2	3	4	5	Average
Teasing	Embarrassment	22.3(23.4)	25.8(27.8)	48.9(23.9)	44.0(26.9)	64.2(23.5)	41.1(17.8)
	Guilt	30.9(24.2)	31.3(23.5)	50.4(24.3)	48.4(27.3)	60.3(26.7)	44.3(18.6)
	Shame	26.8(25.4)	27.8(25.9)	47.8(25.9)	45.8(27.4)	60.3(23.4)	41.7(19.4)
	Average SCEs	26.7(22.7)	28.3(24.6)	49.1(22.7)	46.1(25.5)	61.6(22.6)	42.3(17.4)
	Anger	2.9(9.9)	3.4(10.9)	4.5(12.5)	5.3(15.4)	7.3(18.5)	4.7(12.1)
	Fear	5.4(11.4)	5.9(10.1)	10.0(15.0)	10.6(17.5)	13.7(19.5)	9.1(11.7)
	Sadness	10.2(19.2)	10.8(20.6)	18.1(22.8)	21.7(26.8)	29.1(28.3)	18.0(20.3)
Farting	Embarrassment	18.0(22.5)	17.8(19.7)	36.3(22.3)	47.3(24.7)	62.4(23.6)	36.4(17.8)
	Guilt	10.8(17.4)	9.8(11.7)	16.2(17.5)	17.8(20.4)	25.7(24.8)	16.0(15.9)
	Shame	12.4(20.3)	13.0(15.8)	21.1(21.0)	26.9(24.4)	37.1(28.0)	22.1(19.0)
	Average SCEs	13.7(17.7)	13.5(13.5)	24.5(17.1)	30.7(19.6)	41.7(21.6)	24.8(15.3)
	Anger	2.2(6.4)	1.3(3.9)	3.9(9.4)	4.6(8.9)	8.1(14.5)	4.0(6.9)
	Fear	11.4(19.4)	6.1(11.8)	7.8(16.6)	9.1(17.3)	12.3(21.5)	9.4(14.7)
	Sadness	1.3(3.4)	2.0(6.4)	4.3(8.8)	6.9(13.4)	9.4(17.7)	4.8(8.5)
Dishonesty	Embarrassment	14.2(18.4)	15.5(19.1)	33.1(22.4)	38.1(22.7)	44.5(23.4)	29.1(16.4)
	Guilt	29.9(24.3)	25.4(21.9)	37.0(23.2)	39.6(24.7)	41.6(27.6)	34.7(21.5)
	Shame	19.4(20.1)	21.2(22.4)	31.9(23.0)	34.3(25.1)	40.5(29.5)	29.5(21.2)
	Average SCEs	21.1(18.2)	20.7(18.7)	34.0(21.2)	37.3(21.9)	42.2(25.5)	31.1(18.8)
	Anger	8.8(14.2)	10.4(15.0)	15.5(17.5)	10.5(13.0)	19.0(17.4)	12.8(11.9)
	Fear	21.6(19.9)	13.3(16.1)	24.6(20.2)	23.6(22.4)	27.3(22.8)	22.1(16.6)
	Sadness	6.4(12.6)	5.3(11.5)	7.2(10.7)	11.5(16.8)	11.6(14.8)	8.4(10.1)
Break up	Embarrassment	1.7(6.7)	34.6(26.9)	37.1(27.5)	43.6(30.8)	46.8(31.6)	32.8(19.0)
	Guilt	2.0(10.2)	8.6(16.0)	9.9(17.3)	10.8(17.2)	9.4(17.4)	8.1(12.0)
	Shame	2.9(11.9)	20.9(26.9)	15.9(22.1)	21.6(28.3)	21.8(28.8)	16.6(18.8)
	Average SCEs	2.2(6.9)	21.4(18.6)	21.0(18.2)	25.3(20.5)	26.0(21.1)	19.2(13.8)
	Anger	4.2(14.8)	52.0(27.1)	49.5(26.2)	48.9(28.0)	47.1(28.1)	40.3(20.6)
	Fear	2.5(10.0)	23.7(25.3)	26.9(26.8)	28.4(28.1)	26.1(28.8)	21.5(20.3)
	Sadness	6.2(20.9)	69.3(22.9)	68.2(21.3)	65.6(23.6)	64.5(26.4)	54.8(17.6)
Sabotage	Embarrassment	13.6(18.5)	14.5(20.6)	16.9(20.8)	19.6(25.7)	18.9(27.7)	16.7(19.3)
	Guilt	7.1(14.4)	4.3(9.4)	5.5(11.4)	5.2(11.1)	4.4(10.8)	5.3(9.8)
	Shame	6.2(12.1)	6.3(13.5)	7.0(12.6)	8.6(17.5)	8.8(18.1)	7.4(13.3)
	Average SCEs	9.0(13.2)	8.3(12.8)	9.8(12.7)	11.1(16.6)	10.7(16.9)	9.8(12.9)
	Anger	22.4(27.6)	59.5(23.2)	56.9(21.5)	58.4(23.1)	63.6(22.9)	52.2(19.5)
	Fear	9.2(14.5)	8.3(14.3)	9.0(13.7)	9.6(15.0)	9.5(15.0)	9.1(12.6)
	Sadness	11.4(18.7)	24.5(22.3)	25.1(23.4)	22.4(23.1)	23.1(22.9)	21.3(19.4)

Appendix B: Vignettes Used in Experiment 1

Self-conscious Emotion Scenario 1: Teasing Scenario

Private knowledge 1

You have a good friend Lisa who has a slight speech impediment, always pronouncing "r" as "w." For example, she calls your mutual friend "Wobert" instead of "Robert." You and your friends sometimes joke about Lisa's impediment, but you would never do so when she is around.

One day at the local shopping mall, you run across your friend Robert, and out of habit call out, "Wobert!"

Robert laughs at the inside joke just as Lisa appears out of a nearby store, far enough away that she could not have heard you over the crowd.

Private knowledge 2

You have a good friend Lisa who has a slight speech impediment, always pronouncing "r" as "w." For example, she calls your mutual friend "Wobert" instead of "Robert." You and your friends sometimes joke about Lisa's impediment, but you would never do so when she is around.

One day at the local shopping mall, you run across your friend Robert, and out of habit call out, "Wobert!"

Robert laughs at the inside joke just as Lisa appears out of a nearby store. She hears you make fun of her impediment. But you don't realize this, mistakenly thinking that she is far enough away that she could not have heard you.

Shared knowledge 1

You have a good friend Lisa who has a slight speech impediment, always pronouncing "r" as "w." For example, she calls your mutual friend "Wobert" instead of "Robert." You and your friends sometimes joke about Lisa's impediment, but you would never do so when she is around.

One day at the local shopping mall, you run across your friend Robert, and out of habit call out, "Wobert!"

Robert laughs at the inside joke just as Lisa appears out of a nearby store. She hears you make fun of her impediment and her facial expression changes. You immediately spot this out of the corner of your eye, but you pretend not to have noticed Lisa's reaction, continuing to act as though nothing out of the ordinary has happened. As far as Lisa is concerned, you are unaware that she overheard the joke. When she joins you it is clear that she does not realize that you were aware that she heard you.

Shared knowledge 2

You have a good friend Lisa who has a slight speech impediment, always pronouncing "r" as "w." For example, she calls your mutual friend "Wobert" instead of "Robert." You and your

friends sometimes joke about Lisa's impediment, but you would never do so when she is around.

One day at the local shopping mall, you run across your friend Robert, and out of habit call out, "Wobert!"

Robert laughs at the inside joke just as Lisa appears out of a nearby store. She hears you make fun of her impediment and her facial expression changes. You try not to make eye contact with her by quickly looking in the opposite direction; however, you realize she managed to catch you in the act of looking away. Yet, by this time, you had already turned far enough that it does not appear to her that you noticed that she saw you.

Common knowledge

You have a good friend Lisa who has a slight speech impediment, always pronouncing "r" as "w." For example, she calls your mutual friend "Wobert" instead of "Robert." You and your friends sometimes joke about Lisa's impediment, but you would never do so when she is around.

One day at the local shopping mall, you run across your friend Robert, and out of habit call out, "Wobert!"

Robert laughs at the inside joke just as Lisa appears right behind you out of a nearby store, you all three make eye contact, and it is obvious that she heard the joke.

Self-conscious Emotion Scenario 2: Farting Scenario

Private knowledge 1

While sitting in the middle of a crowded lecture hall, you feel severe bloating in your intestines and immediately regret the bean salad you ate for lunch. You're trapped in the middle of a row with tight seats and do not want to cause a disturbance, but you're in such intense pain that you cannot concentrate on the lecture. Without other options, you decide to try to pass gas silently. You succeed and the fart goes unnoticed.

Private knowledge 2

While sitting in the middle of a crowded lecture hall, you feel severe bloating in your intestines and immediately regret the bean salad you ate for lunch. You're trapped in the middle of a row with tight seats and do not want to cause a disturbance, but you're in such intense pain that you cannot concentrate on the lecture. Without other options, you decide to try to pass gas silently. Unfortunately, the person sitting next to you hears the fart and briefly glances over. But because the room is noisy, and you are focused on the lecture, you don't notice her glance, and mistakenly think you have succeeded in doing so unnoticed.

Shared knowledge 1

While sitting in the middle of a crowded lecture hall, you feel severe bloating in your intestines and immediately regret the bean salad you ate for lunch. You're trapped in the

middle of a row with tight seats and do not want to cause a disturbance, but you're in such intense pain that you cannot concentrate on the lecture. Without other options, you decide to try to pass gas silently. Unfortunately, the person sitting next to you hears the fart and briefly glances over. While you secretly notice her glance, you do not react, and so she has no way of knowing that you noticed; she goes back to taking notes. That is, as far as she knows, you think you have succeeded in passing gas silently.

Shared knowledge 2

While sitting in the middle of a crowded lecture hall, you feel severe bloating in your intestines and immediately regret the bean salad you ate for lunch. You're trapped in the middle of a row with tight seats and do not want to cause a disturbance, but you're in such intense pain that you cannot concentrate on the lecture. Without other options, you decide to try to pass gas silently. Unfortunately, the person sitting next to you hears the fart and briefly glances over. You notice her glance, and reflexively flinch in response, which she sees. However you never actually look over, and so she doesn't think you are aware that she saw you flinch; she goes back to taking notes. That is, as far as she knows, you do not know that she saw you flinch when she heard you.

Common knowledge

While sitting in the middle of a crowded lecture hall, you feel severe bloating in your intestines and immediately regret the bean salad you ate for lunch. You're trapped in the middle of a row with tight seats and do not want to cause a disturbance, but you're in such intense pain that you cannot concentrate on the lecture. Without other options, you decide

to try to pass gas silently. Unfortunately, the person sitting next to you hears the fart and briefly glances over. You can't help but look over and meet her gaze, making it obvious to both you and her that she has noticed the fart.

Self-conscious Emotion Scenario 3: Dishonesty Scenario

Think of a school or community organization that you have been a part of in the past: a sports team, a theater club, a church or political group, or an interest-based group like a ski club. Now imagine that you and your roommate, Sam, both just joined this organization, and you go to the first meeting together.

The first meeting of the year begins with you and Sam both meeting the club members who seem like your kind of people, and you are really looking forward to being a part of the group. Everyone is enjoying small talk over snacks when the President of the group briefly takes the floor to explain the policies and customs of the group. She emphasizes that the group has a tight budget, and that there is a strict policy regarding receipts for club expenses to make sure they are used properly, and so that they are not found in violation of the rules if they are audited by the national organization. She goes on to say it is a very serious infraction to charge personal expenses to the group.

After the announcements, the club selects people for different roles, and you volunteer to get drinks for all of the meetings. You are told to turn in a reimbursement form with your receipts to the Treasurer for reimbursement.

A week later, you and your roommate go to the second meeting with the drinks, and have a great time, but realize when you get home that not only did you forgot to turn in your receipt, but now you can't find it anywhere. You have been on a tight budget yourself, and are stressed about having to absorb the cost for the drinks that you brought to the meeting.

Before the next meeting, you buy some drinks for yourself, which you bring home. Since you were never properly reimbursed at the last meeting, you think it is fair to simply turn in this receipt, since the drinks you bought for yourself were equal to what you bought for the last meeting. You realize that even though this is a perfectly fair solution in terms of cost to the club, it is against the rules. You ask Sam whether you should do this, and Sam says that not only would it be wrong to do so, but that it isn't that much money anyways.

Private knowledge 1

You are annoyed that because of a bureaucratic requirement you are out the money that you are entitled to from the last meeting, so you submit the personal receipt with your reimbursement form. Sam has no idea that you are doing this.

Private knowledge 2

You are annoyed that because of a bureaucratic requirement you are out the money that you are entitled to from the last meeting, so you submit the personal receipt with your reimbursement form. Before submitting the reimbursement form, you leave it on the kitchen table at your house, where Sam sees it and discovers that you are trying to break the rules, but chooses not to do anything about it. However, you were not at home at the time, so you never realize that Sam has seen it.

Shared knowledge 1

You are annoyed that because of a bureaucratic requirement you are out the money that you are entitled to from the last meeting, so you submit the personal receipt with your reimbursement form. Before submitting the reimbursement request, you leave it on the kitchen table at your house, where Sam sees it and discovers that you are trying to break the rules, but chooses not to do anything about it. Peeking out your bedroom door, you see Sam pick up the form and look at it. Sam puts it down without ever realizing you saw him.

Shared knowledge 2

You are annoyed that because of a bureaucratic requirement you are out the money that you are entitled to from the last meeting, so you submit the personal receipt with your reimbursement form. Before submitting the reimbursement request, you leave it on the kitchen table, where Sam sees it. Peeking out your bedroom door, you see Sam pick up the form and look at it. Sam now knows that you were trying to break the rules, but chooses not to do anything about it. However, at that point the door creaks slightly, and as Sam begins to turn to see you, you quickly step back from the door.

Common knowledge

You are annoyed that because of a bureaucratic requirement you are out the money that you are entitled to from the last meeting, so you submit the personal receipt in your reimbursement form. Before submitting the reimbursement request, you leave it on the kitchen table. While Sam has it in hand and is reading it you walk into the room, and before you notice Sam is reading the request, you say hi. Sam looks up at you, you make eye contact, and realize that he has discovered your improper reimbursement form.

Basic Emotion Scenario 1: Break up Scenario

No knowledge

You have been dating your partner for just over 3 years. You are deeply in love with them, think that they are probably "the one", and have been seriously contemplating marriage. You don't live together, but you both have keys to each other's houses and it's not uncommon to be at their house even when they're not home. One afternoon you decide to head over to their house a bit earlier than planned so that you can surprise them with dinner.

You arrive at their house while they're still at work, and are excited because now you know dinner will be a surprise. While you are preparing the meal their phone rings, but you can't get to it in time to answer because you have to wash your hands first. The answering

machine gets the call, and you hear one of their good friends start to leave a message. You decide to let the machine take it at this point, since you would just have to deliver the same message anyways. In the message, the friend says, “Hey, I hope you’re doing OK. Have you decided when you are going to have the breakup talk? I know these things are hard, and so please feel free to reach out to me if you want any support after you end your relationship.” However, because you immediately resumed cooking when the answering machine picked up, you don’t hear any of the message over the running water of the sink.

Private knowledge 1

You have been dating your partner for just over 3 years. You are deeply in love with them, think that they are probably “the one”, and have been seriously contemplating marriage. You don’t live together, but you both have keys to each other’s houses and it’s not uncommon to be at their house even when they’re not home. One afternoon you decide to head over to their house a bit earlier than planned so that you can surprise them with dinner.

You arrive at their house while they’re still at work, and are excited because now you know dinner will be a surprise. While you are preparing the meal their phone rings, but you can’t get to it in time to answer because you have to wash your hands first. The answering machine gets the call, and you hear one of their good friends start to leave a message. You decide to let the machine take it at this point, since you would just have to deliver the same message anyways. The answering machine is loud and right by the kitchen, so you can hear the entire message. The friend says, “Hey, I hope you’re doing OK. Have you decided when

you are going to have the breakup talk? I know these things are hard, and so please feel free to reach out to me if you want any support after you end your relationship." You don't know what to do, so you pack up the food you brought over and head home.

Private knowledge 2

You have been dating your partner for just over 3 years. You are deeply in love with them, think that they are probably "the one", and have been seriously contemplating marriage. You don't live together, but you both have keys to each other's houses and it's not uncommon to be at their house even when they're not home. One afternoon you decide to head over to their house a bit earlier than planned so that you can surprise them with dinner.

You arrive at their house while they're still at work, and are excited because now you know dinner will be a surprise. While you are preparing the meal their phone rings, but you can't get to it in time to answer because you have to wash your hands first. The answering machine gets the call, and you hear one of their good friends start to leave a message. You decide to let the machine take it at this point, since you would just have to deliver the same message anyways. The answering machine is loud and right by the kitchen, so you can hear the entire message. The friend says, "Hey, I hope you're doing OK. Have you decided when you are going to have the breakup talk? I know these things are hard, and so please feel free to reach out to me if you want any support after you end your relationship." You don't know what to do, so you pack up the food you brought over and head home. Unbeknownst

to you however, your partner sees you pull out of their driveway, and so they will realize that you heard the message.

Shared knowledge

You have been dating your partner for just over 3 years. You are deeply in love with them, think that they are probably "the one", and have been seriously contemplating marriage. You don't live together, but you both have keys to each other's houses and it's not uncommon to be at their house even when they're not home. One afternoon you decide to head over to their house a bit earlier than planned so that you can surprise them with dinner.

You arrive at their house while they're still at work, and are excited because now you know dinner will be a surprise. While you are preparing the meal their phone rings, but you can't get to it in time to answer because you have to wash your hands first. The answering machine gets the call, and you hear one of their good friends start to leave a message. You decide to let the machine take it at this point, since you would just have to deliver the same message anyways. The answering machine is loud and right by the kitchen, so you can hear the entire message. The friend says, "Hey, I hope you're doing OK. Have you decided when you are going to have the breakup talk? I know these things are hard, and so please feel free to reach out to me if you want any support after you end your relationship." You don't know what to do, so you pack up the food you brought over and head home. However, you see your partner driving down the street as you pull out of the driveway, so you know that your partner will realize that you heard the message.

Common knowledge

You have been dating your partner for just over 3 years. You are deeply in love with them, think that they are probably "the one", and have been seriously contemplating marriage. You don't live together, but you both have keys to each other's houses and it's not uncommon to be at their house even when they're not home. One afternoon you decide to head over to their house a bit earlier than planned so that you can surprise them with dinner.

You arrive at their house while they're still at work, and are excited because now you know dinner will be a surprise. While you are preparing the meal their phone rings, but you can't get to it in time to answer because you have to wash your hands first. The answering machine gets the call, and you hear one of their good friends start to leave a message. You decide to let the machine take it at this point, since you would just have to deliver the same message anyways. Right then your partner walks in, makes eye contact with you, and you both hear the friend's message say, "Hey, I hope you're doing OK. Have you decided when you are going to have the breakup talk? I know these things are hard, and so please feel free to reach out to me if you want any support after you end your relationship."

Basic Emotion Scenario 2: Sabotage Scenario

No knowledge

You go to the mall to buy some much-needed new clothes. After making your purchase you walk out of the store and see two of your friends, Mike and John, hanging out and talking. Mike is one of your best friends, but you and John aren't especially close, and you're only really friends because you are both in the same friend group. You and John get along fine, he just isn't someone who you would normally hang out with if he wasn't always hanging out with your other friends.

They don't see you walking up, and you realize they are talking about you when you overhear one of them say your name. John is lying about you to Mike saying that you always say bad things about him when he's not around, and that you told him you don't really like Mike very much, and follows that up with, "I just don't understand how you could be friends with such a jerk." However, even though you were close enough to hear your name, you weren't actually able to hear anything else that John said, so you don't realize he was bad-mouthing you to Mike. Before either of them realizes that you are there, you duck into a nearby store.

Private knowledge 1

You go to the mall to buy some much-needed new clothes. After making your purchase you walk out of the store and see two of your friends, Mike and John, hanging out and talking. Mike is one of your best friends, but you and John aren't especially close, and you're only really friends because you are both in the same friend group. You and John get along fine, he just isn't someone who you would normally hang out with if he wasn't always hanging out with your other friends.

They don't see you walking up, and you realize they are talking about you when you overhear John say your name. Before either of them notices that you are within hearing distance, you hear John lying about you to Mike saying that you always say bad things about him when he's not around, and that you told him you don't really like Mike very much. It's apparent that he's trying to sabotage your friendship, since he follows this up with, "I just don't understand how you could be friends with such a jerk." Before either of them realizes that you overheard them, you duck into a nearby store.

Private knowledge 2

You go to the mall to buy some much needed new clothes. After making your purchase you walk out of the store and see two of your friends, Mike and John, hanging out and talking. Mike is one of your best friends, but you and John aren't especially close, and you're only really friends because you are both in the same friend group. You and John get along fine, he just isn't someone who you would normally hang out with if he wasn't always hanging out with your other friends.

They don't see you walking up, and you realize they are talking about you when you overhear John say your name. Before either of them notices that you are within hearing distance, you hear John lying about you to Mike saying that you always say bad things about him when he's not around, and that you told him you don't really like Mike very much. It's apparent that he's trying to sabotage your friendship, since he follows this up with, "I just don't understand how you could be friends with such a jerk." You don't think

that either of them has noticed you yet, and you duck into a nearby store. However, unbeknownst to you John actually saw you duck into the store and realizes you must have heard him.

Shared knowledge

You go to the mall to buy some much-needed new clothes. After making your purchase you walk out of the store and see two of your friends, Mike and John, hanging out and talking. Mike is one of your best friends, but you and John aren't especially close, and you're only really friends because you are both in the same friend group. You and John get along fine, he just isn't someone who you would normally hang out with if he wasn't always hanging out with your other friends.

They don't see you walking up, and you realize they are talking about you when you overhear John say your name. Before either of them notices that you are within hearing distance, you hear John lying about you to Mike saying that you always say bad things about him when he's not around, and that you told him you don't really like Mike very much. It's apparent that he's trying to sabotage your friendship, since he follows this up with, "I just don't understand how you could be friends with such a jerk." You don't think that either of them has noticed you yet and you duck into a nearby store. However, right as you duck into the store you look back and see John turn his head back towards Mike, and you realize that he must know you heard him.

Common knowledge

You go to the mall to buy some much needed new clothes. After making your purchase you walk out of the store and see two of your friends, Mike and John, hanging out and talking. Mike is one of your best friends, but you and John aren't especially close, and you're only really friends because you are both in the same friend group. You and John get along fine, he just isn't someone who you would normally hang out with if he wasn't always hanging out with your other friends.

They don't see you walking up, and you realize they are talking about you when you overhear John say your name. Before either of them notices that you are within hearing distance, you hear John lying about you to Mike saying that you always say bad things about him when he's not around, and that you told him you don't really like Mike very much. It's apparent that he's trying to sabotage your friendship, since he follows this up with, "I just don't understand how you could be friends with such a jerk." Just as you are about to duck into a nearby store, John looks over at you and you make eye contact with him, making it obvious to both of you that you heard what he said.

Appendix C: Full Instructions Given to Participants in Experiment 2

Common Knowledge Condition

Recently, many reality TV shows have been introduced in which a panel of judges judge someone's performance. However, the relationship between the judges and the performers may impact their judgments. In this study we're interested in how judges' anonymity may affect their judgments.

For each study we recruit a panel of judges and a performer. In this experiment you have been assigned the role of the performer, and I have already set the judges up in the room across the hallway. You will perform a karaoke song, which they will observe over a live two-way video feed. We are using the video feed so that we can use the same environment both when the judges are anonymous and when they are not. In one condition the judges know that you know they are there, and in the other they believe they are anonymous and that you don't know they are there.

You have been selected for the condition in which the judges know that you are aware of their presence. You will be able to see them on the two-way video feed during your performance, and they know that you can see them through the two-way video feed. In other words, both you and the judges know that you are watching each other over a two-way video feed during your performance.

Is this all clear, and do you have any questions?

After the song ends, the feed will cut off, and you will click on a button to answer a few questions in a very short survey about your performance. Once you finish the survey, your part in the experiment is done, and you can come out of the room and let me know you're finished.

Finally, in order to keep your data anonymous, you have been assigned a Participant ID number that will not be associated with your name. I have written it down on this post it note that I will leave with you in the room, because you will have to enter this number at the beginning of the survey. Please make sure you enter the number correctly, because it is the only way we will be able to identify your data.

Any questions?

Alright, then let's begin. You can just follow me into this room over here.

Shared Knowledge Condition

Recently, many reality TV shows have been introduced in which a panel of judges judge someone's performance. However, the relationship between the judges and the performers may impact their judgments. In this study we're interested in how judges' anonymity may affect their judgments.

For each study we recruit a panel of judges and a performer. In this experiment you have been assigned the role of the performer, and I have already set the judges up in the room across the hallway. You will perform a karaoke song, which they will observe over a live two-way video feed. We are using the video feed so that we can use the same environment both when the judges are anonymous and when they are not. In one condition the judges know that you know they are there, and in the other they believe they are anonymous and that you don't know they are there.

You have been selected for the anonymous condition, in which the judges think that you are unaware of their presence. However, in order to keep the performances the same across conditions, we actually always tell the performers about the judges. In addition, you will be able to see them on the two-way video feed during your performance, but they don't know this. In other words, even though you can see the judges over this feed, they think that you don't know they are there, and that they are simply watching a one-way video feed of your performance.

Is this all clear, and do you have any questions?

After the song ends, the feed will cut off, and you will click on a button to answer a few questions in a very short survey about your performance. Once you finish the survey, your part in the experiment is done, and you can come out of the room and let me know you're finished.

Finally, in order to keep your data anonymous, you have been assigned a Participant ID number that will not be associated with your name. I have written it down on this post it note that I will leave with you in the room, because you will have to enter this number at the beginning of the survey. Please make sure you enter the number correctly, because it is the only way we will be able to identify your data.

Any questions?

Alright, then let's begin. You can just follow me into this room over here.